Structure and Dynamics of the *E. coli* Chaperone, Trigger Factor, and its Interaction with Ribosome-Nascent Chain Complexes

Anne S. Wentink

Research Department of Structural and Molecular Biology,
University College London

2014

A dissertation submitted for the degree of
Doctor of Philosophy in Structural and Molecular Biology

Declaration

I, **Anne S. Wentink**, declare that all the work presented in this thesis is the result of my work only. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. The work herein was carried out while I was a graduate student at the University College London, Research Department of Structural and Molecular Biology, under the supervision of Prof John Christodoulou.

Abstract

The *de novo* folding of newly synthesised polypeptide chains, exposed in a vectorial manner to the crowded cellular environment, often requires the assistance of molecular chaperones. The ribosome-associated molecular chaperone, trigger factor (TF), facilitates the early folding events of such nascent chains at the ribosomal exit tunnel. The investigation of the structural and dynamic properties of this interaction, at high resolution, presents a challenge due to the dynamic nature of the interaction and the intrinsic conformational heterogeneity of nascent chains. Additional complexity comes from the competition between TF self-association into a dimeric state, the interaction with ribosome-nascent chain complexes (RNCs) and the interaction with isolated protein substrates independent of the ribosome.

The sensitivity of NMR spectroscopy to both structural and dynamic changes make this technique uniquely suited to the investigation of the trigger factor chaperone. This thesis presents the application of selective isotopic labelling strategies and advanced TROSY optimised NMR experiments to the study of the TF dimerisation equilibrium, providing a detailed understanding of its kinetics and thermodynamics and insights into the solution structure of the TF dimer. The improved understanding of the mechanism of TF dimerisation forms the basis for the interpretation of observations made in the presence of substrate proteins.

Preliminary results on the interaction of TF with two disordered model ribosome-nascent chain complexes reveal that ribosome-associated nascent chain substrates are significantly perturbed by the presence of TF, even in the absence of predicted TF binding motifs. The study of TF and the corresponding isolated polypeptides reveals that limited interaction occurs in the absence of the ribosome. This early NMR investigation of the TF chaperone provides the first residue-specific details of the highly dynamic interaction of TF with nascent polypeptides and begins to elucidate the role of this chaperone in the process of co-translational folding.

Acknowledgements

The work presented in this thesis was carried out at the Department of Structural and Molecular Biology, University College London, UK over the course of 4 years. It was driven by the thoughts, collaboration and encouragements of many.

First of all, I would like to express my gratitude to my supervisor, Professor John Christodoulou for his support and guidance every step of the way, from biochemistry undergraduate through to completion of this degree. His honest council has always been invaluable.

I am indebted to the BBSRC funding body which has supported me financially during my PhD studies, and to Professor Gabriel Waksman and Dr Carolyn Moores for giving me the opportunity to undertake my doctoral studies as a member of the ISMB. I sincerely thank Prof Waksman and Dr Andrew Osborne for guiding me through my PhD unscatched as members of my thesis committee. Furthermore, I'm grateful to my scientific collaborators, Dr Enrico Salvadori and Prof Chris Kay, for their contribution to the EPR aspects of the present study.

I also wish to thank the past and present members of the JC group as well as Dr John Kirkpatrick for creating a great dynamic environment which has been a joy to work in. I'm particularly grateful to Dr Lisa Cabrita and Dr Marilia Karyadi for welcoming me into the group as an undergraduate student and showing me the ropes. I further want to thank Dr Christopher Waudby for the time spent teaching me NMR and his keen insight and mathematical prowess. I would also like to thank my fellow PhD students/ coffee addicts for their support and camaraderie. I want to thank in particular Anaïs Caissaignau and Annika Weise, whom I've worked with closely in the preparation of this work, for their friendship, feedback and kind words of encouragement.

I want to finish off by thanking my family and Sebastian Geibel who have never failed to support me. Finally, I want to dedicate this thesis to my brother, Mark Wentink - it's your turn now.

Contents

Abbreviations

aa	amino acids
ADP	adenosine diphosphate
Amp	Ampicillin
APS	ammonium persulfate
ATP	adenosine triphosphate
a.u.	arbitrary units
BME	β-mercaptoethanol
CFTR	cystic fibrosis transmembrane conductance regulator
CI2	chymotrypsin inhibitor 2
CPMG RD	Carr-Purcell Meiboom-Gill relaxation dispersion
cryo-EM	cryo-electron microscopy
CSA	chemical shift anisotropy
CSP	chemical shift perturbation
D. discoideum	*Dictyostelium discoideum*
ddFLN	*Dictyostelium discoideum* filamin
ddFLN5 (ddFLN6)	domain 5 (6)
DHFR	dihydrofolate reductase
DNA	deoxyribonucleic acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
E. coli	*Escherichia coli*
EDTA	ethylenediamine-tetraacetic acid
EM9	enhanced M9 medium
EPR	electron paramagnetic resonance
FID	free induction decay
FKBP	FK506-binding protein
FRET	Förster resonance energy transfer
FT	Fourier transform
GFP	green fluorescent protein
HbS	sickle-cell haemoglobin

HDX	hydrogen deuterium exchange
HMQC	heteronuclear multiple-quantum coherence
HSQC	heteronuclear single-quantum coherence
ICDH	isocitrate dehydrgenase
IDP	intrinsically disordered protein
IPTG	isopropyl-β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
LB	Luria-Bertani medium
MAP	methionine aminopeptidase
MBP	maltose binding protein
MDG	Phosphate (M), L-aspartic acid (D) and glucose (G) based medium
mRNA	messenger RNA
MW	molecular weight
MWCO	molecular weight cut-off
NAC	nascent-chain associated complex
NC	nascent chain
Ni-IDA	Nickel-Iminodiacetic acid
Ni-NTA	Nickel-NitriloTriacetic acid
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NS	number of scans
O/N	overnight
OD	optical density
PCR	polymerase chain reaction
PDF	peptide deformylase
PCS	pseudo-contact shifts
PhoA	alkaline phosphatase precursor
PPIase	peptidyl-prolys *cis* / *trans* isomerase
ppm	parts per million
PRE	paramagnetic relaxation enhancement
PTC	peptidyl-transferase centre
RBD	ribosome-binding domain
RDC	residual dipolar couplings
RNA	ribonucleic acid
RNase A	ribonuclease A
RNC	ribosome-nascent chain complex
SRP	signal recognition particle
rpm	revolutions per minute

RT	room temperature
S	Svedberg (non-SI sedimentation unit)
SBD	substrate-binding domain
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SecM	secretion monitor
sHSP	small heat-shock protein
SOFAST	band-selective optimised-flip-angle short-transient
SORDID	Signal Optimisation with Recovery in Diffusion Delays
STE	stimulated echo
TBS	Tris-buffered saline
TEV	tobacco Etch virus
TF	trigger factor
Tico	"tight-couple" (buffer that permits the extraction of 70S ribosomes i.e. "tight-couple" 30S and 50S as described in [1])
TMS	tetramethylsilane
tRNA	transfer RNA
TROSY	transverse relaxation optimised spectroscopy

Chapter 1

Introduction

1.1 Protein folding

1.1.1 *In vitro* protein folding

The correct folding of a polypeptide chain is essential to achieving full protein functionality. The 'native' folded state typically corresponds to a narrow ensemble of structurally similar conformations [2], while fluctuations herein are likely to be essential for protein function [3]. Such biologically important fluctuations in the protein structure include conformational changes associated with substrate binding as described by the *induced fit* principle [4]) and intra-cellular communication by means of molecular switch mechanisms, as seen for example in G-protein coupled receptors which undergo large conformational changes in response to an extracellular signal [5].

To ensure these structural changes can occur, the native three-dimensional conformation is only marginally stable [6]. The origin of the marginal stability of a protein's native fold can be found in the competing thermodynamic forces that drive folding (Fig 1.1, [7]) described by the equation for Gibbs' free energy: $\Delta G = \Delta H - T\Delta S$. Indeed, protein folding results in the formation of numerous weak, non covalent interactions decreasing the system's enthalpy (ΔH). This decreased free energy of the system is however offset by an abrupt loss in entropy (ΔS) upon chain compaction, leading to an overall energetic stability of on average only 10 kcal/mol [8, 9]. Fluctuations within the structural ensemble required for function are typically associated with energies not much greater than the thermal energy (k_bT, where k_b is the Bolzmann constant) of the system, permitting their occurrence on nanosecond timescales [3].

Our current understanding of the forces that drive protein folding have been obtained by means of detailed *in vitro* denaturing and renaturing studies of small proteins [2, 15]. These studies identified the gain in H_2O entropy upon burial of hydrophobic residues, the 'hydrophobic effect', as the principle driving force of protein folding, while further

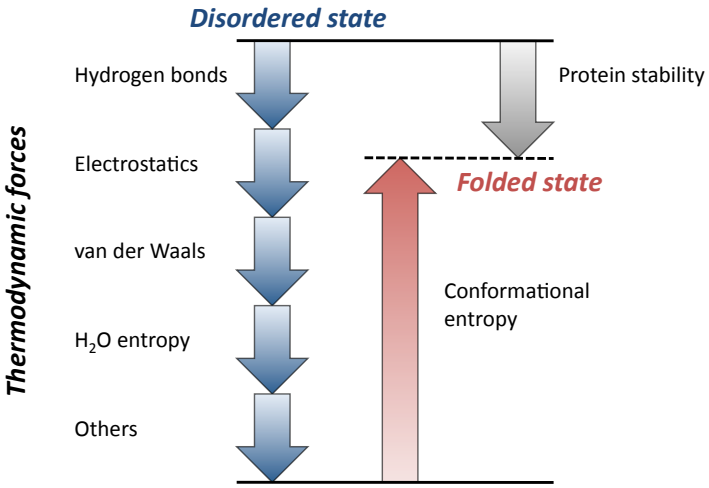

Figure 1.1: Thermodynamic forces drive protein folding. Protein folding is driven by both enthalpic forces such as hydrogen-bond networks, favourable electrostatic interactions and van der Waals forces as well as entropic forces in the form of the hydrophobic effect. Folding is opposed, in large part, by the significant reduction of the polypeptide chain's conformational entropy upon folding. The net force that drives protein folding is thus typically small resulting in marginally stable native protein states.

stabilising forces include the increased enthalpy from the formation of hydrogen bonds, electrostatic and van der Waals interactions as well as disulfide bridge formation, summarised in Fig 1.1 [6]. Folding is opposed by a decrease in conformational entropy upon compaction of the polypeptide chain. It is the competition between these constructive and disruptive entropic forces that leads to folding barriers and potential kinetically trapped, 'intermediate' folding states. These 'kinetic traps' typically arise when long range stabilising contacts need to be broken to rearrange to a more native structure [16].

Based on specific folding events observed in single protein studies, models for the folding pathways of proteins have been proposed to integrate the various driving forces (Fig 1.2). The various models conflict on the order of secondary and tertiary structure formation and prove difficult to reconcile into an unified thermodynamical description of all studied folding events. At one extreme it has been proposed in a model referred to as 'diffusion-collision' that secondary structural elements transiently form independently before colliding into the final tertiary structure [13]. According to this model, protein folding requires the transition through an intermediate state where secondary structure is formed in the absence of tertiary contacts. Such a short lived species was found on the folding pathway of barnase through kinetic experiments [10]. NMR experiments found that this intermediate species presented fully formed α-helical and β-sheet elements prior to full structure acquisition [17]. At the other extreme, 'hydrophobic collapse' or condensation has been suggested as the initial event in folding resulting in an unstructured but collapsed

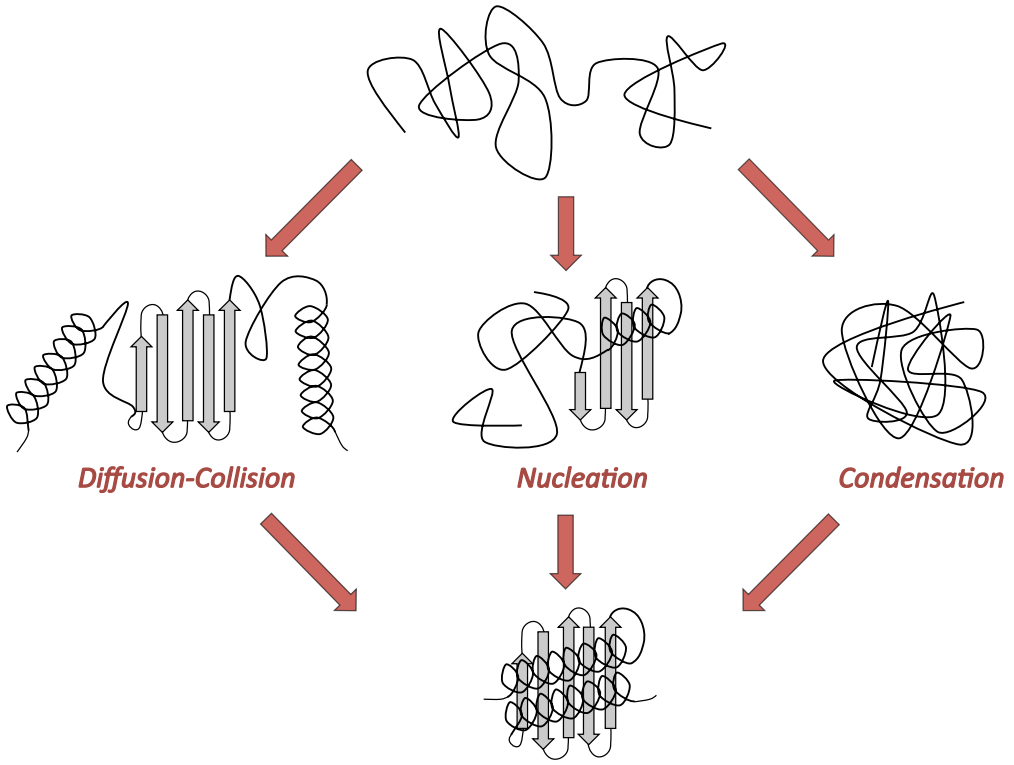

Figure 1.2: Classical view of protein folding pathways. Proposed models of protein folding pathways based on detailed characterisation of the folding mechanisms of proteins such as barnase [10], chymotrypsin inhibitor 2 [11] and lysozyme [12] (refer to text) differing in the order of secondary and tertiary structure formation. In the diffusion-collision model (left), secondary structural elements form and then collide to make tertiary contacts [13]. In the nucleation model (centre), secondary and tertiary contacts propagate from one initial secondary structural element [14]. In the condensation model (right), the structure collapses to a molten-globule state where tertiary contacts can begin to form before any secondary structure appears [12].

state where constructive tertiary contacts are formed before entire secondary structural elements develop. A particular example where such a collapsed state was observed is the folding of lysozyme and α-lactalbumin (reviewed in [12]). This 'molten globule' state presents a rudimentary hydrophobic core within which stable secondary structural elements can form. Slow rearrangements of these elements around the collapsed core lead to the native folded state. Finally a theory that sits at the interface of these two views relies on the nucleation of a single structural element which then propagates along the sequence forming stabilising tertiary contacts along the way (reviewed in [14, 18]). This theory has been used to explain simple two-state folding phenomena such as observed for chymotrypsin inhibitor 2 (CI2) where folding does not transition through a stable intermediate species. Instead, the formation of the single α-helix in the native structure leads to a rapid 'global collapse' of the structure [11]. In conclusion, the three classical views of

protein folding were developed independently on the back of specific protein models but are not capable of reconciling all observed folding phenomena

An overarching theory has been proposed as a unified way of studying protein folding that sees the afore described folding pathways as specific cases of the same basic folding principles. The energy landscape theory of protein folding proposes that spontaneous protein folding follows a steep funnel-like trajectory from a fully unfolded, random coil conformation to a lowest energy state [20] as illustrated in Figure 1.3.a. Although the final fold is defined, the pathway followed can vary dramatically because of the large conformational space of the unfolded protein. Both the final state and the available pathways down the funnel are believed to be defined primarily by the protein amino acid sequence composition [7]. Local minima in this 'rugged' energy landscape sampled by some, but not all accessible folding pathways correspond to trapped intermediate folding states [21, 22] which are functionally inactive and can be particularly aggregation prone due to the incomplete burial of hydrophobic core residues [23]. At the bottom of the constructive pathway(s), active and inactive intermediates are separated by small energy barriers allowing functionally important fluctuations in the three-dimensional structure [3].

Misfolding events are in direct competition with productive folding pathways and can lead to the formation of highly stable oligomeric species that are prone to aggregation. Although particular proteins present much higher propensity to misfold [24], amyloid formation appears to be a generic property of proteins based on a inherent propensity of the polypeptide chain to form β-sheet hydrogen bonding networks. Protein aggregation and amyloid formation is driven by the same forces involved in native protein folding where the formation of extensive hydrogen bond networks and the burial of hydrophobic residues in extensive quaternary interfaces represent the majority. It is thus the kinetic competition between the formation of constructive and aggregation interactions that biases a polypeptide chain down a particular side of the folding funnel (Fig 1.3.a).

The importance of correct folding in the cell is illustrated by numerous disease phenotypes strongly correlated to the misfolding of proteins. Common neurodegenerative disorders such as Parkinson's, Alzheimer's disease are caused by aggregation of α-synuclein and Aβ and Tau respectively [25, 26] while the aggregation prone haemoglobin mutant HbS lays at the origin of sickle-cell anaemia [27]. Correct folding of the polypeptide chain is thus paramount to all cellular processes and the maintenance of a correct folding balance is one of the most prioritised functions in the cell.

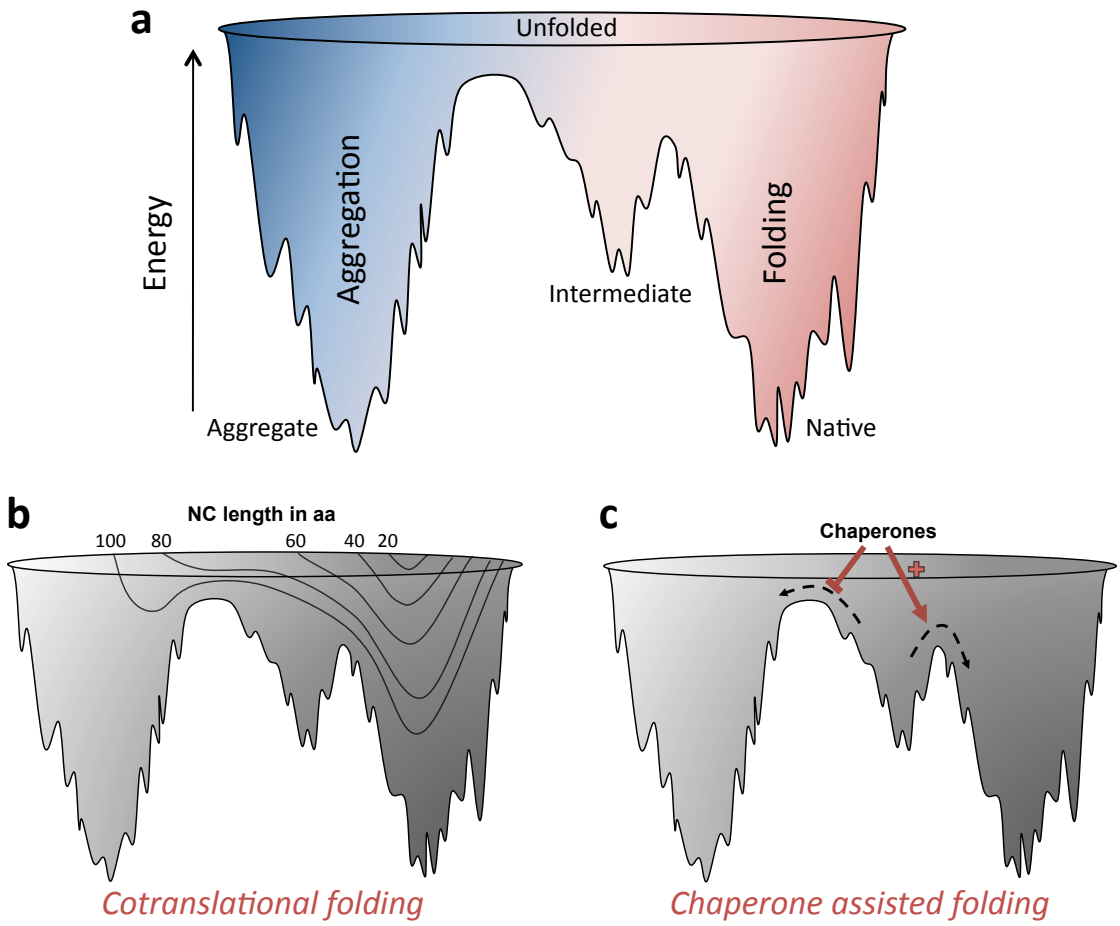

Figure 1.3: Energy landscape of protein folding and aggregation. (**a**) Folding funnel displaying the competition between folding and aggregation. Both pathway are separated by a large energy barrier. Polypeptides can become trapped in inactive intermediate states which are particularly aggregation-prone. (**b**) Co-translational folding alters the available configurational space as more of the primary sequence information becomes available (Section 1.1.3); increasing lengths of polypeptides can explore a larger part of the energy landscape. Co-translational folding is believed to have an inherent bias for constructive folding (Section 1.1.3). (**c**) Chaperones can prevent the formation of aggregated species or promote the passage of an intermediate species onto a productive folding pathway (Section 1.2). Adapted from [19].

1.1.2 Protein folding in the cell

The environment of the cell however differs greatly from *in vitro* conditions. First of all, the cellular environment is highly crowded with protein concentrations reaching 400 mg/ml [28]. This greatly increases the risk of non-productive intermolecular protein interactions [23] and aggregation [29] but might also be conducive to rapid folding due to a bias for collapsed, compact conformations [30, 31]. Cellular crowding is likely to greatly increase the energy barrier between aggregation and folding as well as increasing the ruggedness of the funnel appearance (Fig 1.3) corresponding to an increased stability of, off-pathway, intermediate structures which would require large conformational rearrangements for recovery. Moreover, cellular crowding strongly affects diffusion rates [32]. As a consequence, long range contact formation (encounter time) is often the rate limiting step in protein folding in the cell corroborated by the observation that the folding rate of a protein strongly correlates with the contact order of the 'native' state: the average distance in primary sequence between tertiary contacts [33, 34].

Protein synthesis in the cell occurs in a vectorial manner on the ribosome. The nascent chain thus has the opportunity to form intramolecular contacts and interact with other cytosolic factors in the absence of full primary sequence information. Protein folding can thus be initiated in a co-translational manner. The differences between co-translational folding and the refolding of full-length proteins *in vitro* are the subject of current investigations further described Section 1.1.3.

Furthermore, the presence of potential interaction partners that are absent within the controlled *in vitro* environment can significantly affect protein folding pathways. For instance, proteins such as Fibronectin Binding Protein A undergo significant structural changes upon interaction with binding partners (Fibronectin) [35]. Other proteins rely on the presence of cofactors to stabilise their functional fold and might require these to nucleate their initial folding (e.g. zinc-finger domains which rely on the coordination of four cysteines to a zinc ion to induce loop-type structure which is essential for DNA binding [36]).

Finally, the difficulty of refolding multi-domain proteins *in vitro* [37] suggests other factors within the cell are required for the effective independent folding of structural units within larger structures. These factors include both temporal control by regulation of the speed of translation by rare codon clustering [38, 39] or the prevention of non-native inter-domain contacts by chaperones [40].

The extent to which *in vitro* and *in vivo* folding pathways differ is the subject of current studies. In particular, since the energy changes associated with folding are small, minor perturbations in the landscape through the factors listed above, could severely impact the folding pathway and stabilities of intermediates. Reconciling our theoretical

knowledge based on extensive *in vitro* studies with observations of folding events in the cellular environment is one of the great challenges in the field of protein folding.

1.1.3 Co-translational protein folding

1.1.3.1 The ribosome

Protein biosynthesis in all domains of life is assured by large ribonucleoprotein complexes named ribosomes (Fig 1.4). They are responsible for the decoding of genetic information and the translation of this information to specific amino acid sequences. The small 30S subunit binds mRNA and recruits the large 50S subunit which harbours the site of peptide bond formation. A complex set of concerted motions which have been elucidated to exquisite detail by extensive structural study, result in the decoding of the mRNA via reverse complementation of the sequence by specific recruited tRNA-amino acid complexes [41, 42]. The correct alignment of two consecutive tRNA molecules results in peptide bond formation between the two amino acids. A set of motions subsequently displaces the initial tRNA and allows for the recruitment of tRNA to the next codon in sequence, and further addition of amino acids to the nascent polypeptide chain.

The growing polypeptide chain is displaced from the peptidyl-transferase centre (PTC) through a long 100 Å exit tunnel that traverses most of the 50S subunit.The exit tunnel is estimated to accommodate around 30 amino acid in extended conformation. The bacterial ribosomal exit tunnel is lined by the 23S rRNA and ribosomal proteins L4, L22 and L23 and has a diameter of 10-20 Å [47]. The exit tunnel widens in the last 20 Å to what is called the 'exit vestibule' where nearby ribosomal proteins L22, L23, L24 and L29 form docking sites for ribosome-associated factors such as peptide deformylalse (PDF), methionine aminopeptidase (MAP), trigger factor (TF) and signal recognition particle (SRP) involved in the processing, targeting and folding of nascent chains in *E. coli*, Fig 1.4 [48].

1.1.3.2 Evidence for co-translational folding

The vectorial synthesis of proteins on the ribosome results in a unique scenario of protein folding. The sequential addition of amino acids at the PTC results in the sampling of conformational space by the N-terminal residues in the absence of full primary sequence information [49, 50]. The available conformational space thus expands in a length dependent manner [19](Fig 1.3.b) and might bias constructive folding pathways over misfolding events by preventing the accumulation of non constructive intermediate species. Research over the last decade described below indicate that the nascent polypeptide can begin to sample native-like conformations while attached on the ribosome [51]. Folding events during translation, also called co-translational folding, thus appear to defy the existing

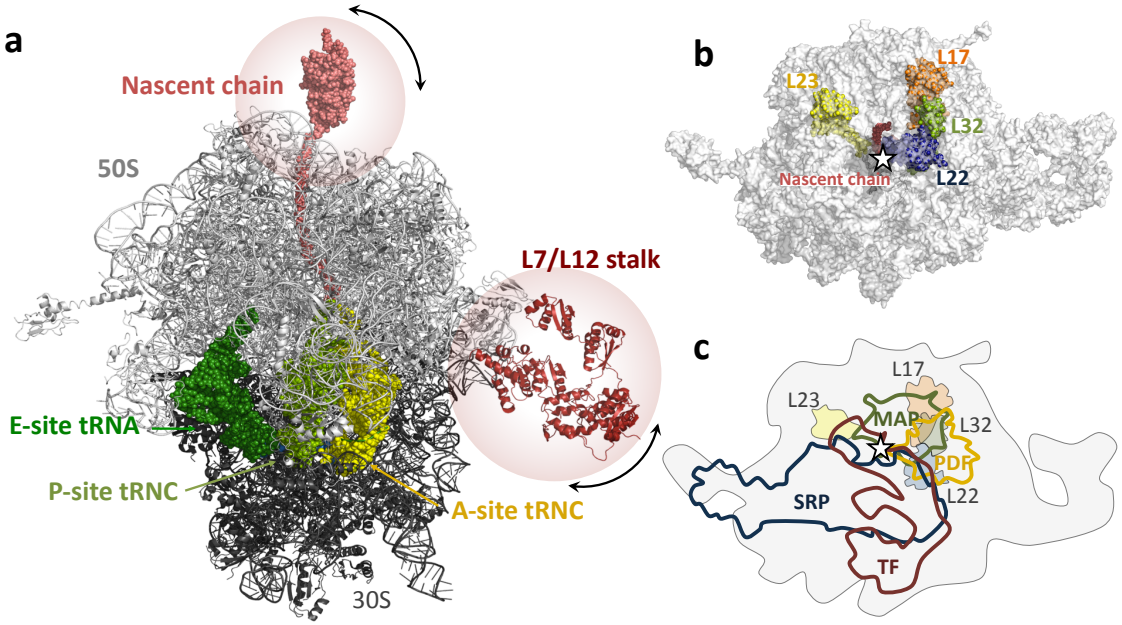

Figure 1.4: Ribosome structure and interaction with ribosome-associated factors. (**a**) Structural model of the 70S ribosome complex in the presence of a ddFLN5 nascent chain. The interface between the 30S (dark grey, PDB 2J00 [43]) and 50S (light grey, PDB 2J01 [43]) accommodates the mRNA decoding activity. The A-, P-, and E-site tRNAs are shown in yellow, green and dark green respectively (PDB 2J00, A-site tRNA modelled from PDB 1GIX [44]). The petidyl-transferase centre (PTC) is located in the 50S at the extremity of the A- and P-site tRNAs where the ribosomal exit tunnel starts. A nascent chain (pink, poly-A sequence modelled in exit tunnel attached to the globular ddFLN5 domain, PDB 1QFH [45]) traverses the 100 Å exit tunnel before it can acquire both structure and flexibility. In red is highlighted the flexible L7/L12 stalk region (modelled from PDB 1RQU [46]) highlighting the potential similarities in the dynamics of both these regions. (**b**) Top view of 70S ribosome particle. Ribosomal proteins L17 (orange), L22 (blue), L23 (yellow) and L32 (green) near the exit vestibule serve as docking sites for ribosome-associated factors. (**c**) Schematic representation of ribosome-associated factors trigger factor (TF, red), signal recognition particle (SRP, blue), methionine aminopeptidase (MAP, green) and peptide deformylase (PDF, yellow) as bound on the ribosome near the exit tunnel (marked by star symbol).

paradigm that the whole primary sequence of a protein is required for native-like structure formation.

Hints of secondary structure have been observed by cryo-electron microscopy (cryo-EM) within the ribosomal exit tunnel [52, 53] for specific protein sequences while pegylation [54] and Förster resonance energy transfer (FRET) studies [55] have observed similar compaction in discreet regions of the ribosomal exit tunnel. α-helical conformations appear to be able to form in the middle regions of the exit tunnel, but a constriction in the exit tunnel width leading to the exit vestibule formed by L4 and L22 makes it unlikely that this structure persists [56]. In the exit vestibule, significantly more conformational space can be sampled by the nascent chain [47]. Recent cross-linking studies have indicated that

small structural elements such as α and β hairpins can start to form in the exit vestibule [57]. Selective proteolysis of nascent chains found the exit tunnel accommodates 30-40 residues [58, 59] indicating the nascent chain is neither fully extended nor fully helical in the ribosomal exit tunnel.

Sequence dependent stalling of translation (e.g via residues 150-166 of SecM) suggest inter- and intramolecular interactions can occur within the ribosomal tunnel [60–62]. The SecM stalling motif (FxxxxWIxxxxGIRAGP) makes a number of key interactions with ribosomal protein L4 and L22 and the 23S rRNA resulting in a distortion of interactions at the PTC preventing peptide-bond formation which leads to stable arrest of translation at Pro 166 [56]. Similarly, the presence of arginine residues has been shown to significantly reduce translation speed by altering the local electrostatic environment of the ribosomal tunnel [63]. The stalling properties of the 17 amino acid sequence from SecM has been exploited for the *in vivo* production of stalled ribosome-nascent chain complexes (RNCs) permitting equilibrium studies of nascent chains of specific lengths [56, 64, 65]. Although translation is a dynamic process, the rapid folding of small protein domains (<1 s) relative to the rate of translation in particular in eukaryotes where the rate of translation is 3-8 a.a per second [66], suggests that these equilibrium studies may provide an accurate picture of the available conformational space in the absence of slower folding events such a proline isomerisation [67].

Native folding, however, requires the emergence of the entire polypeptide sequence in order to complete final long range contacts [68] in most proteins. Long range contacts have been shown to form close to the ribosomal surface where interaction with ribosomal proteins and in particular the negatively charged rRNA are likely to occur [69]. The formation of native structures on the ribosome has been shown using a variety of techniques including monitoring of native disulfide bridge formation [64, 70], the appearance of enzymatic activity [64, 69, 71, 72], protease protection assays [73, 74] and recognition by conformation specific antibodies [74]. Notable difference in protein folding on the ribosome relative to *in vitro* were observed for T4 lysozyme nascent chains where the ribosome appeared to delay folding [75] while specific on-pathway cotranslational folding intermediates were identified in GFP-like fluorescent protein nascent chains [76].

Co-translational protein folding appears to be particularly important to multi-domain proteins such a luciferase which rely on the absence of subsequent domains to prevent stable non-productive long range contacts from forming [73, 74, 77]. Sequential folding of domains has further been shown for hRas-DHFR (dihydrofolate reductase) and CFTR (cystic fibrosis transmembrane conductance regulator) using limited proteolysis [78, 79] and FRET [80] approaches.

1.2 Molecular chaperones

Protein homeostasis (or proteostasis) within the cell is maintained by a large family of proteins known as molecular chaperones. Their role as modulators of protein folding states makes them essential to a cell's well-being under stress and to general housekeeping functions. Though neither sequence nor structural similarity persists throughout this family, they are grouped by a functional definition of their class as proteins affecting the three-dimensional structure of substrate proteins without pertaining to the physiologically active structure of these substrates [81].

1.2.1 Cellular functions

Within the cell, molecular chaperones are responsible for a variety of functions including the *de novo* folding [82] and re-folding of proteins, the unfolding of proteins targeted for degradation, de-aggregating misfolded proteins [83] and the sequestering of proteins awaiting complex assembly [84, 85]. The cellular functions of molecular chaperones in *E. coli* are summarised in Fig 1.5 and Table 1.1.

Ubiquitous and highly versatile chaperones are those interacting with their substrates co-translationally or soon after translation to initiate *de novo* folding (Fig 1.5). In *E. coli* major representatives of this group are trigger factor (TF), DnaJ/K and GroEL/ES, and together, this chaperone network ascertains the *de novo* folding of emerging nascent chains on the ribosome [103]. These non substrate-specific chaperones are characterised by a high affinity for non buried hydrophobic side chains which is detected as a sign of incorrect folding. DnaJ/K and GroEL/ES actively promote folding through an ATP-driven binding and release cycle [104, 105]. While chaperone binding prevents aggregation in the crowded cellular environment (Fig 1.3 C), release is needed to permit correct burial of hydrophobic residues in the final protein fold. Chaperones involved in *de novo* protein folding demonstrate a significant redundancy in function highlighting the importance of ensuring correct folding of new polypeptide chains.

A specific case of chaperones involved in the *de novo* folding process of emerging nascent chains are chaperones such as SecB involved in the translocation of proteins across the cellular membrane (Fig 1.5) [106]. General transport of proteins by the SecYEG translocon machinery typically occurs in an unfolded state and molecular chaperones are required to maintain this translocation competent unfolded state and target the polypeptide to the membrane, while preventing aggregation in the cytosol. Here, the interaction with a specific highly hydrophobic signal sequence, recognised by chaperones and cytosolic factors such as SRP, prevents the untimely formation of native structural contacts. Within the periplasm, parallel chaperone systems have evolved to support the effective folding of nascent polypeptide in this highly oxidising environment (Fig 1.5) [107].

Table 1.1: Overview of molecular chaperones in *E. coli*

Chaperone	Co-factor	Function	Review
Trigger factor		ribosome-bound chaperone	[86]
DnaK	DnaJ/GrpE	holdase, unfolding	[87]
GroEL	GroES	ATP-dependent folding	[87]
PPIases (YtfC, SlyD, ...)		prolyl isomerase	[88]
SecB		holdase, recruitment to SecYEG	[89]
YidC		membrane insertion	[90]
Skp		periplasmic chaperone	[91]
SurA		periplasmic chaperone	[91]
LolA		lipoprotein chaperone	[92]
Hsp 31		heat-inducible holdase	[93]
Hsp 33		redox-regulated holdase	[94]
HdeA/HdeB		acid survival, periplasmic chaperone	[95]
IbpA/IbpB		holdase	[96]
HtpG		*de novo* folding	[97]
ClpB	DnaK/J/GrpE	ATP-driven de-aggregation	[98]
ClpA/C/X		ATP-dependent unfolding	[99]
DegP		periplasmic folding and degradation	[100]
Pilin chaperones (FimC, PapD, ...)		chaperone-usher pathway	[101]
HscA	HscB	iron-sulfur cluster protein assembly	[102]

ATP-independent chaperones including the family of small Heat Shock Proteins (sHSPs) do not actively drive folding but function as 'holdases' preventing aggregation of misfolded species by sequestering exposed hydrophobic patches [108] in particular under stress conditions (Fig 1.5). Chaperones with ATP-dependent unfolding activity, such as the Hsp100 family [109], are responsible for the de-aggregation of misfolded proteins [110–112] and are often found as integral part of proteolytic pathways [113]. By contrast, substrate specific chaperones have evolved for roles based on sequestering or 'structural inactivation' of proteins awaiting assembly into larger macromolecular complexes including for example the family of chaperone-usher pathway chaperones involved in pilus biogenesis [114].

The eukaryotic cell presents similar chaperone networks ensuring essential protein homeostasis [87]. Additional complexity to the highly redundant networks relative to bacteria originates from further specialisation of homologues based on compartment localisation or tissue specific requirements. Overall, the necessity to preserve correct protein

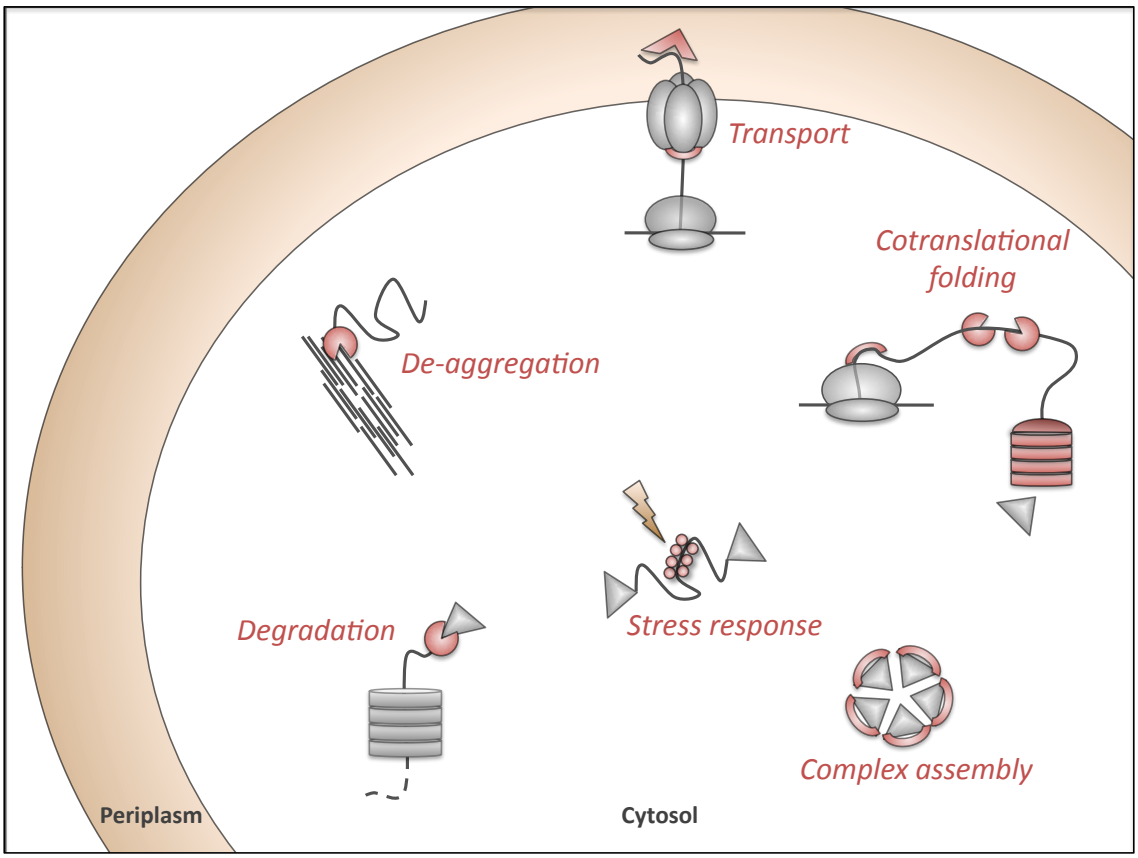

Figure 1.5: Cellular functions performed by molecular chaperones. Molecular chaperones (red) are responsible for a large number of cellular functions related to the maintenance of protein fold homeostasis within the cell. Chaperones perform functions at all stages of a protein's lifetime from initial biogenesis to degradation by the proteosome.

folding states is conserved across all kingdoms and is a highly prioritised function within the cell with some of the most abundant proteins belonging to the class of molecular chaperones [115].

In addition to ensuring the general maintenance of correct folding states within the cell, chaperones might have in part evolved as a compensation mechanism for the accumulation of destabilising mutations on the path to evolution of new protein functions and specificities, while circumventing the negative selective pressure on structurally unstable proteins [116, 117].

1.2.2 Trigger Factor

The ATP-independent molecular chaperone trigger factor (TF) is a highly conserved ribosome-associated chaperone in bacteria. Forming an elaborate network with further molecular chaperones DnaJ/K and GroEL/ES, it is responsible for the correct *de novo*

folding of the proteome [103]. Although knock-out studies of TF show no clear phenotype, double knock-out of TF and DnaK results in a heat-sensitive phenotype marked by protein misfolding and aggregation in *E.coli* [118–122]. The apparent redundancy inherent in this chaperone network reports on the vitality of its function to the cell's survival.

The origin of TFs name can be traced back to its first identification as a protein factor implicated in triggering pro-OmpA insertion into the membrane. Trigger factor was rediscovered in 1995 as a ribosome-associated protein with peptidyl-prolyl cis/trans isomerase (PPIase) activity [123] thus establishing its chaperoning role. The eukaryotic functional equivalent of trigger factor, as first protein to interact with nascent polypeptides [124] is the nascent-chain associated complex (NAC) despite not sharing common structural features [125].

1.2.2.1 Structural organisation

The 48 kDa trigger factor folds into three distinct domains [133, 134] adopting an extended three-dimensional structure frequently likened to a dragon as shown in Fig 1.6 [126]. The N-terminal domain of TF (residues 1-111) presents a mixed α-helical and β-sheet structure related to the bacterial chaperone Hsp33 [135]. It is this domain that mediates the essential interaction with the ribosome through a characteristic ribosome-binding motif (GFRxGxxP) in a helix-loop-helix element absent from the Hsp33 structure [136, 137].

The ribosome-binding domain (RBD) is linked to the second domain by means of a long extended α-helix bridging 65 Å. This peptidyl-prolyl isomerase domain (PPIase) extends from residue 149 to 245 and belongs to the family of FK506-binding protein (FKBP) PPIases [134]. Its catalytic activity is dependent on residues F198, F233, Y221 with the F198A mutation shown to abolish complete PPIase activity [138] and significantly accelerates proline isomerisation in substrates such as RNase T1 [139]. The TF PPIase activity is non essential to the main chaperone function of TF [138]. It has however been suggested to provide an auxiliary binding site for substrates and was shown to increase TF residency on nascent chains [140, 141].

The C-terminal domain (246-432) nests between the PPIase domain and the RBD where it adopts an α-helix rich structure closely related to the periplasmic chaperone SurA, resembling two arms protruding away from the rigid linker helix [126]. This C-terminal domains (CTD) conveys the main chaperoning activity of TF [142, 143] and is in large part responsible for substrate interaction in the absence of the ribosome [84, 130–132, 144].

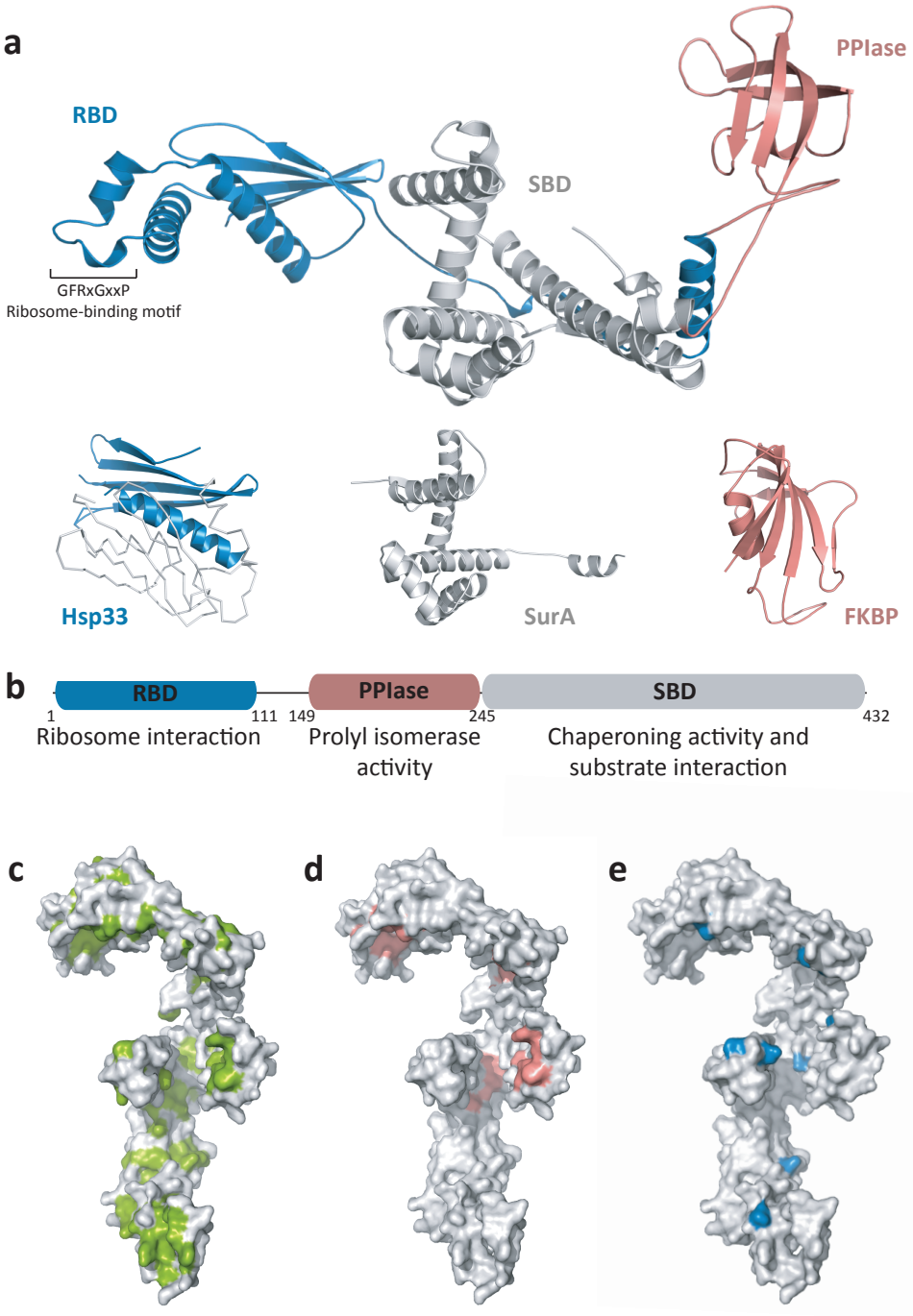

Figure 1.6: Structural organisation of trigger factor. (**a**) The three-domain structure of the *E. coli* trigger factor (PDB 1W26 [126]). The ribosome-binding domain (RBD, blue) shows close homology to parts of the Hsp33 chaperone (PDB 1I7F [127]). The central chaperoning and substrate binding domain (SBD, grey) structurally closely resembles the periplasmic chaperone SurA (PDB 1M5Y [128]). The peptidyl-prolyl isomerase (PPIase) domain belongs to the family of FKBP PPIases (PDB 4IQ2 [129]). (**b**) The structural domain organisation does not follow the linear amino acid sequence. Colours as in a. (**c**) TF structure rotated 90° relative to **a**. Hydrophobic residues (Ala, Leu, Ile, Phe, Trp and Val) are marked in green. (**d**) Substrate interaction sites identified in [130] in red. (**e**) Nascent chain sites of cross-linking identified in [131] and [132] in blue.

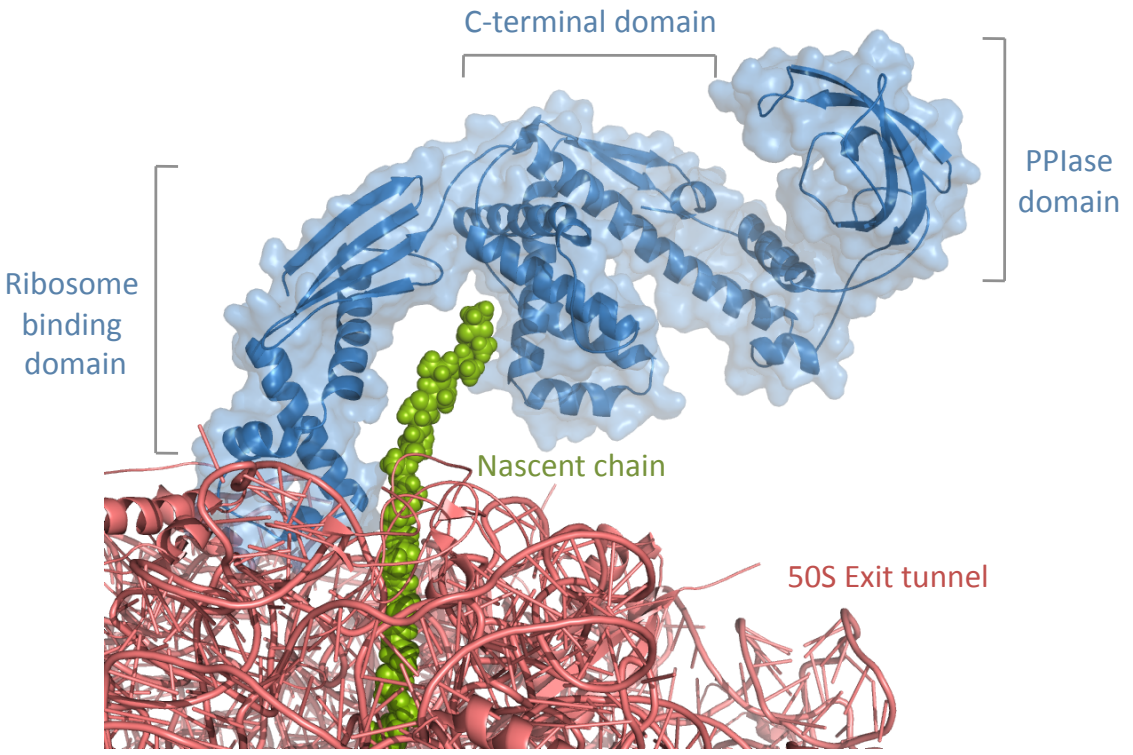

Figure 1.7: Structural model of TF on the ribosome. The *E. coli* trigger factor (PDB 1W26 [126]) modelled onto the structural model of a nascent chain occupied 50S ribosomal subunit (PDB 2J01 [43] and PDB 1QFH [45], see Fig 1.4) based on a cryo-EM model (PDB 2VRH [132]).

1.2.2.2 The trigger factor is a ribosome associated molecular chaperone

Unlike other known molecular chaperones, TF associates with the 70S ribosome in a 1:1 stoichiometry and binds at the intersection of ribosomal proteins L23 and L29 and domain III of the 23S rRNA near the exit tunnel via a conserved ribosome-binding motif [126, 137, 145–147]. The interaction is mediated primarily through electrostatic contacts and is thus salt-sensitive [145] and results in extension of the overall TF structure [148] and specific structural changes in the TF RBD, exposing a hydrophobic surface towards the ribosomal exit vestibule. The interface between ribosomal protein L23 and TF is small, explaining the relative weak affinity of TF for the inactive ribosome (see Section 1.2.2.5).

No high resolution structure of the full-length TF on the ribosome is available to date. Co-crystallisation of the N-terminal RBD with the ribosomal 50S sub-unit allowed for modelling of the complete TF structure on the ribosome [126, 146, 147] (Fig 1.7). This image closely resembles the lower resolution electron microscopy (EM) data of ribosome-bound TF in the presence of a nascent polypeptide chain [132]. Through the RBD interaction near the exit tunnel, the TF CTD ‘arms’ point across the exit tunnel towards

the ribosome surface exposing a large hydrophobic crevice towards the exit tunnel. The sheltered space created underneath TF is surprisingly large and is estimated to be capable of accommodating a globular protein domain of up to 15 kDa [126].

1.2.2.3 Mechanisms of TF interaction with substrate proteins

TF is said to be a promiscuous binder lacking high substrate specificity [84]. However, its interactome has been shown to be enriched in hydrophobic and multi-domain proteins. Selective ribosome profiling has found that in particular large outer membrane β-barrel proteins (the Omp family) are important TF targets [149]. Furthermore, it was found that TF preferentially binds to substrates of more than 100 amino acids in length suggesting competition with early processing factors such as PDF and MAP delays initial binding of TF to the NC in the cell [149].

Peptide screening of 10 amino acid long sequences from *E.coli* proteins indicated that a combination of high hydrophobicity (in particular enrichment in aromatic residues) and positively charged residues form the TF interaction sequence [150]. A number of studies have reinforced the predictive power of these specific characteristics in linear $\sim$10 amino acid stretches to identify TF interaction sites [130, 148].

A recent study, using NMR spectroscopy, established the first solution structure of multiple TF molecules bound to a disordered protein substrate, alkaline phosphatase precursor (PhoA) in the absence of the ribosome, identifying four distinct substrate binding sites [130]. Three locate to the SBD while the last corresponds to the catalytic site of the PPIase domain (Fig 1.8). The substrate interactions are hydrophobic in nature with substrate binding site enriched in hydrophobic and aromatic residues while peripheral charged residues appear to stabilise interactions further through salt-bridge formation. While the four identified substrate binding sites shares these common features their exact characteristic differ, displaying a clear preference for certain regions of the PhoA substrate. The PPIase mediated interaction is most hydrophobic in nature and the TF substrate binding site largely consists of aromatic residues (Fig 1.8.b). The 'neck' and arm 2 binding sites of the SBD are lined primarily with long side chain hydrophilic amino acids with key hydrogen bonds stabilising the interaction (Fig 1.8.c and d). The central cavity binding site of the SBD is more mixed in nature combining both hydrophobic and polar contacts for interaction (Fig 1.8.a). In contrast, the structure of a 2:2 complex of TF and folded ribosomal protein S7 showed extensive hydrophilic interactions covering the entirety of the RBD and CTD cavity [84] (Fig 6.1, page 226).

Cross-linking experiments have identified the TF-nascent chain interaction surface on the ribosome as the entire interior of the TF hydrophobic cradle [132] between the two protruding arm regions (Fig 1.6.e). The nascent-chain has been suggested to progress

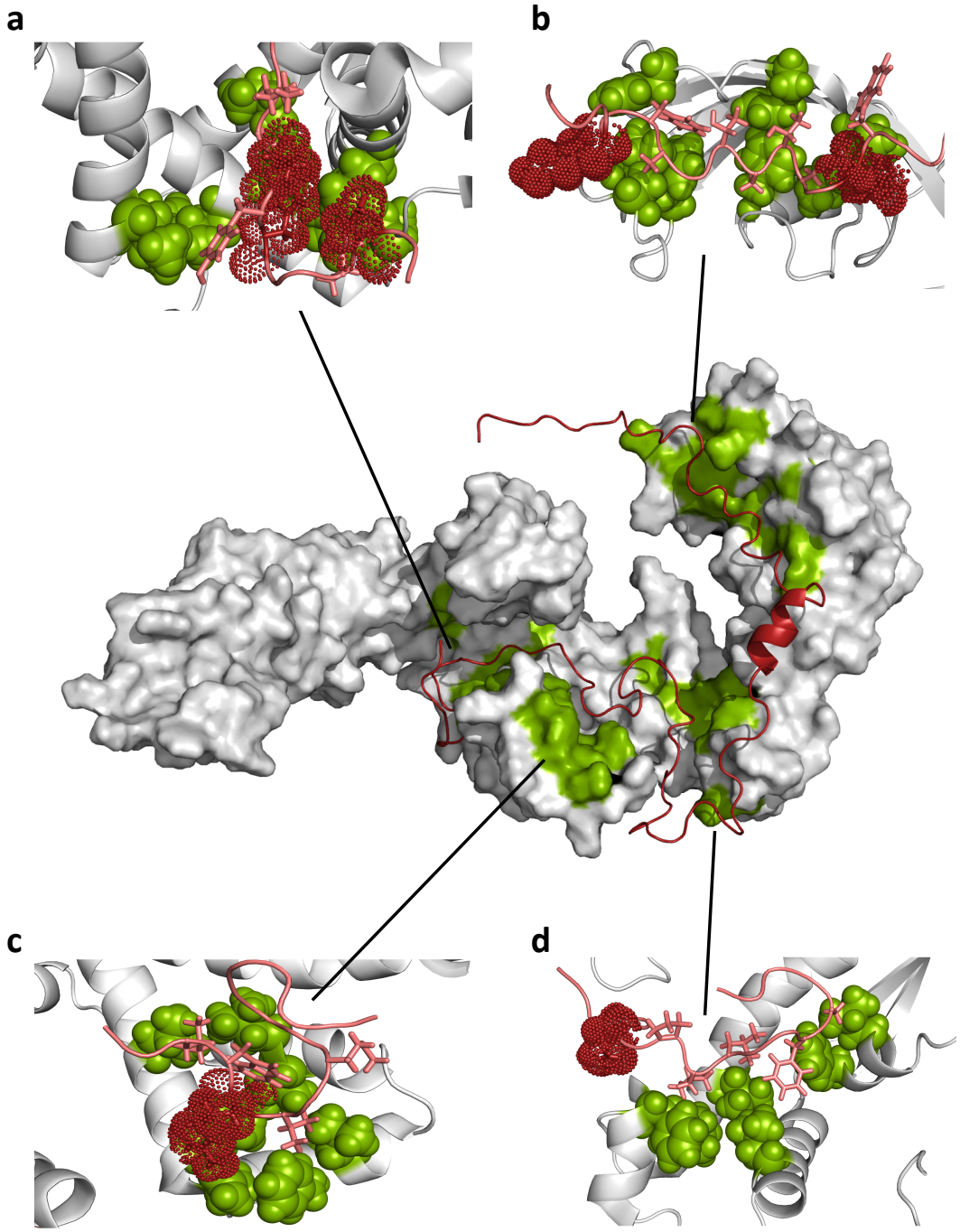

Figure 1.8: NMR structure of monomeric TF in complex with PhoA$_{220-310}$. Lowest energy NMR structure of TF (grey) in complex with PhoA (red) (pdb 2MLX, [130]) with the four substrate interaction site identified in [130] in green. (**a** to **d**) Detail of the four identified substrate binding sites of TF (cavity, PPIase, neck and arm 2 regions respectively). Hydrophobic TF residues involved in substrate interaction are shown in green spheres and the PhoA substrate in pink with residues involved in TF contact in stick representations. PhoA residues involved in polar contacts with TF are shown in red dot representation.

in a length-dependent manner from the exit tunnel between the arm domains, extending into the unprotected cytosolic space while bent over the PPIase 'head'. At an undefined threshold length, the isocitrate dehydrogenase (ICDH) nascent chain appears to retract back under TF presumably adopting a more compact folded structure. TF appears to shields disordered and some folded nascent chains (up to 46 kDa) [121, 132, 151–153].

1.2.2.4 Functional models of TF interaction with substrates

Despite extensive structural and biochemical investigations the exact mechanism of the TF chaperoning action still lacks clarity. Recent studies have proposed several distinct roles for TF (Fig 1.9). Initial models of foldase activity were based on a combination of the above cross-linking studies which traced the length-dependent interactions of TF with the ICDH nascent chain and the cryo-EM and crystallographic images of TF crouched over the ribosomal exit tunnel [126, 132] It was suggested that TF aids the active folding of nascent chains by restricting conformational space and increasing local concentration of potential intra-chain contacts Fig 1.9 [126]. Moreover, the highly charged nature of the ribosome surface and the mixed nature of the *E. coli* TF cradle lead to the suggestion that TF would drive the formation of a hydrophobic core through repulsion in a Anfinsen cage-type mechanism [84].

Recently, a strong case has been built for a model attributing a 'holdase' function to TF (Fig 1.9). Interactions with hydrophobic NCs would prevent non-native contacts before the entire primary sequence of a domain becomes available. These models are based on the observation that TF preferentially binds hydrophobic stretches in protein chains [148] remaining bound until the protein spontaneously folds and buries these patches thus breaking the TF-NC complex. TF was found to stabilise on pathway folding intermediates in maltose binding protein (MBP) in a recent study using optical tweezers [154]. This model is further supported by studies showing that TF can remain bound to NCs while no longer attached to the ribosome [84] such that multiple TF molecule can accumulate on the nascent chain until a domain becomes folding competent. Moreover, TF was found to delay folding in multi-domain proteins firefly luciferase and β-galactosidase [140, 155]. The holdase model of TF action is further corroborated by the structure of TF in complex with the unfolded PhoA where the winding of the substrate between the various binding site results in a fairly extended conformation [130]. This would also permit the sequential folding of domains in multi-domain proteins while blocking interfering interactions with the subsequent domain. TF was also found to unfold folding-competent substrates such as barnase and MBP [64, 130]. Unclear is whether TF achieves this through stabilisation of transient unfolded species or through active unfolding of the folded state (Fig 1.9).

Other chaperoning functions have been proposed independently from its immediate

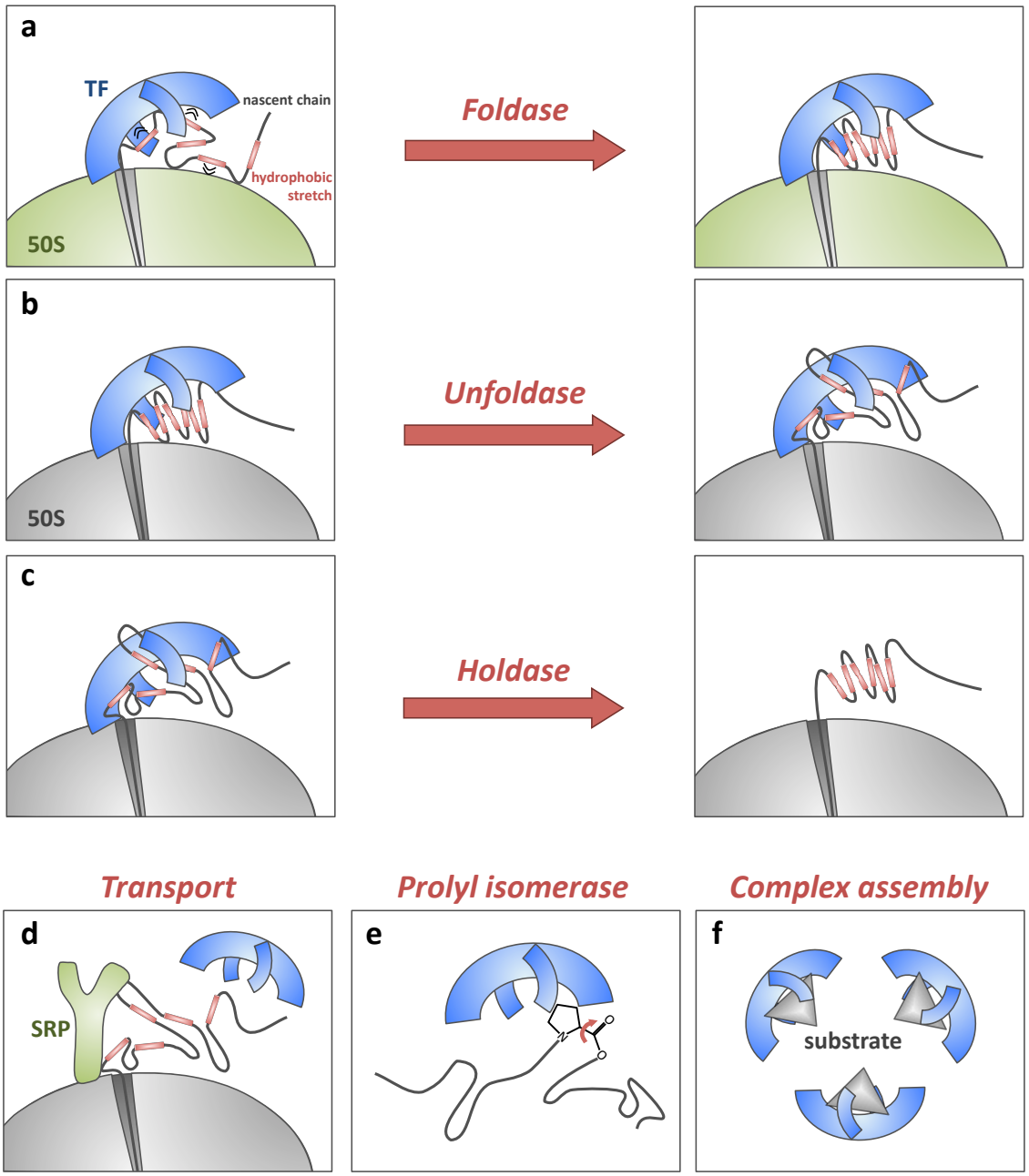

Figure 1.9: Proposed cellular functions of the molecular chaperone trigger factor. Based on structural data, the cage like appearance of TF at the ribosomal exit tunnel let to the proposal of (**a**) an Anfinsen cage based chaperoning mechanism where TF and the ribosome surface (green) form a hydrophilic cage, driving the burial of hydrophobic residues in the nascent chain. (**b** and **c**) Delayed folding and the destabilisation of folding competent nascent chains suggest TF possesses 'unfoldase' and 'holdase' activity preventing folding till complete folding units are synthesised. (**d**) TF was further implicated in the regulation of translation of nascent chains through competition with SRP. (**e**) TF's PPIase domain activity is dependent on substrate interaction of the C-terminal chaperoning domain. (**f**) Identification of folded TF substrates, in particular ribosomal proteins, have suggested a role for TF in complex assembly through sequestration of aggregation-prone complex subunits.

effect on the nascent-chain's tertiary structure. Among them is a potential role in co-translational translocation by competition for substrates with other NC interacting proteins such as Signal Recognition Particle (SRP) [147, 156] and SecB (Fig 1.9) [157]. Recently, co-crystallisation of a TF dimer with ribosomal protein S7 after a proteomics approach had shown most ribosomal proteins are relative high affinity (low μM) TF substrates, has resulted in the proposal of a role for TF in ribosome assembly [84] (Fig 1.9).

Finally the importance and role of the PPIase domain remains elusive. Its distance away from the ribosomal exit tunnel provides a potential explanation for its proven minor contribution to TF chaperoning activity as demonstrated by knock-out and deletion studies [140, 141] although it effectively accelerates the folding of proline-limited substrates [133, 139, 158]. In most models, while the TF hydrophobic cavity permits initial folding, the PPIase domain can follow up quickly on non-native, exposed proline residues. Independently, the additional binding site at the PPIase catalytic centre permits the stabilisation of the TF-NC complex [131, 138, 152]. The central position of the PPIase domain in the primary sequence would argue it could possibly have been maintained as an artefact of the evolutionary fusion of a ribosome-binding protein and a SurA-like chaperone.

While the chaperoning activity of TF resides in the SBD, the PPIase domain has a significant impact on protein folding in it own right because of its catalytic *cis*/*trans* peptidyl-prolyl isomerase activity. It is likely that TF combines many of the described mechanisms favouring one or the other on a substrate-specific basis. The difficulty to reconcile observations of the TF action on specific substrates hints at the versatility of TF to adapt to the specific need of a substrate protein. This ability to adapt appears to be a common feature of molecular chaperones with many fulfilling different roles dependent on the associated co-factor [87].

1.2.2.5 Trigger factor equilibria

To facilitate its various functions, TF exists in a fine-tuned, highly dynamic equilibrium between various substrate bound states as summarised in Fig 1.10. In the absence of substrates, TF exist in a monomer-dimer equilibrium with an dissociation constant of ~1-20 μM [148, 159]. Although high affinity folded protein substrates exist [84] with dissociation constants in the low μM range, the majority of non-ribosome associated TF will adopt this dimeric form which has been suggested to serve as a storage form where occlusion of the TF substrate interface prevents unwanted interactions in the cytosol.

The monomeric form of TF binds the ribosome at L23 in a 1:1 stoichiometry with a 1-2 μM affinity [148, 159, 160]. TF's hydrophobic interior [38, 83, 84] provides a non specific interface for nascent chains with exposed hydrophobic patches. Unsurprisingly, TF preferentially binds to highly hydrophobic nascent chain substrates (Kd ~1-700 nM)

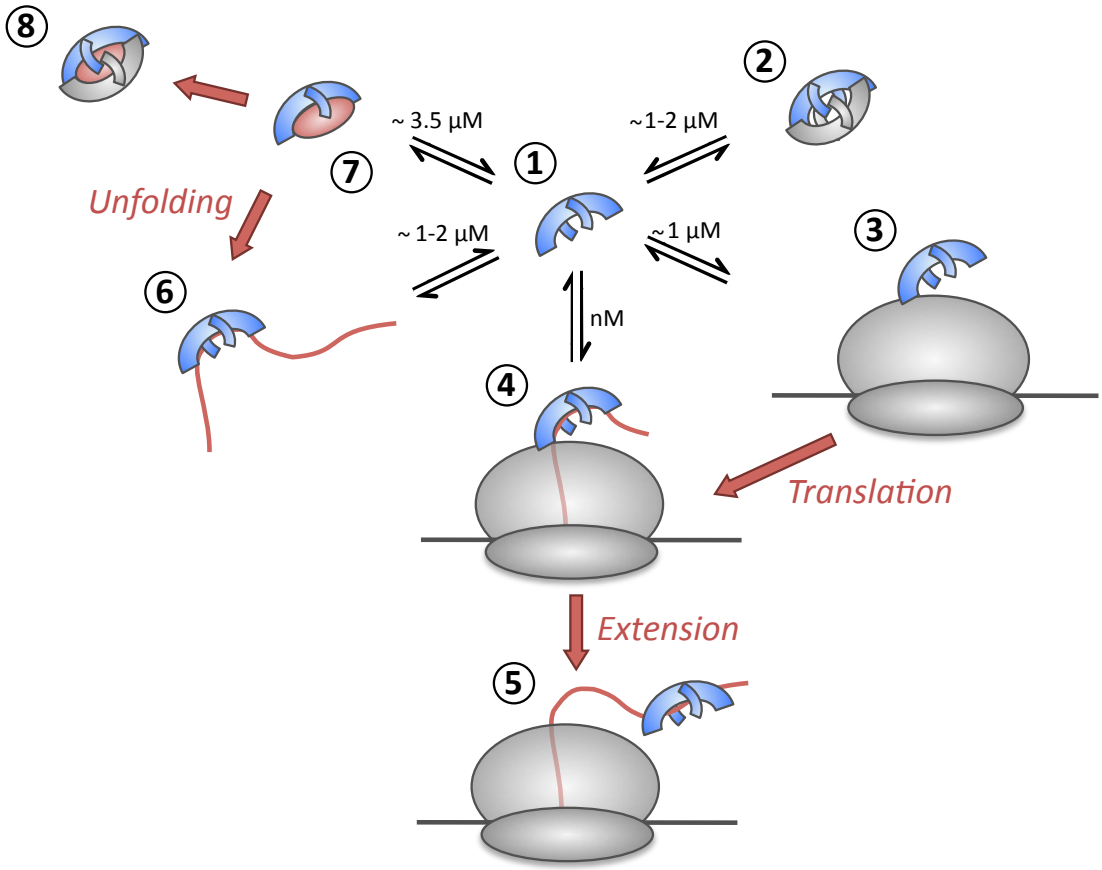

Figure 1.10: Model of TF equilibrium within the cell based on experimentally identified TF interactions. TF (blue) exist in a dynamic equilibrium between various substrate (red) bound states. The monomer (**1**) self-associates with μM affinity into a dimeric form (**2**) effectively burying it's interaction surface. With similar affinity, TF transiently interacts with inactive ribosomes (**3**) while preferring nascent chain occupied ribosomes with nM affinity (**4**). TF can remain associated with nascent chains after dissociation from the ribosome (**5**). TF further interacts with low affinity with some unfolded (**6**) and folded (**7**) protein substrates independently from the ribosome, interactions which have been suggested might be important for complex assembly.

ensuring that TF is recruited to those nascent chains that benefit most from the protection it offers [85]. The residency time of TF on the ribosome is highly reliant on the nature of the nascent chain produced. Residence time on empty ribosomes is estimated to be ~10-15 seconds allowing screening of ribosomes for nascent chain occupancy. Nascent chains containing hydrophobic stretches can extend the residency time of TF up to 50 seconds [148, 161]. The exact values for the TF-RNC half-lives are currently being challenged with a recent study indicating that the kinetics of TF-70S and TF-RNC interaction are much more rapid than previously measured. Specific affinities for the ribosome and select RNCs were also found to be tighter than previously determined at 100 nM and 1 nM respectively

[162].

TF may remain associated with emerging nascent chains when dissociated from the ribosome [155]. This allows the recruitment of multiple TF molecules to the same nascent chain or the interaction of down-stream regions with other ribosome-binding factors such as SRP. Both folding of cytosolic domains by TF and membrane insertion of transmembrane regions by SRP/SecYEG of an inner membrane protein could thus take place simultaneously.

The interaction of TF with substrates off the ribosome is in direct competition with dimerisation and therefore occurs with relatively high affinity substrates only (1-5 μM). While interaction with unfolded substrates is highly transient, TF was shown to form stable complexes with folded domains such as S7 [84]. A more in-depth discussion of the TF equilibria can be found Section 3.4.

To conclude, the molecular chaperone TF is the paradigm of ribosome-associated chaperones. Its structure in isolation and function on the ribosome have been investigated extensively and have shed light on a highly diverse and highly adaptable molecule. While TF is known to undergo conformational changes upon interaction with the ribosome, no high resolution structure of TF bound to ribosome is available to date. Furthermore, a solid understanding of the dynamic interplay between the various substrate interactions discussed above and the conformational changes associated with each state remains elusive.

1.3 Nuclear magnetic resonance spectroscopy

Proteins undergo highly dynamic processes such as folding, which cannot be defined by a single conformational state. The sensitivity of nuclear magnetic resonance (NMR) spectroscopy to both structural and dynamic transitions make it the method of choice for the high resolution study of structural rearrangements and dynamic differences during protein folding. The origin of the NMR signal and an overview of the phenomena underlying most of the structural and dynamic NMR experiments used in this study are detailed further in Section 2.1. In the following, previous applications of NMR to the study of dynamic systems and in particular large macromolecular complexes such as the ribosome are presented as examples of the versatility of NMR spectroscopy.

1.3.1 NMR and dynamics

As described in Section 1.1, proteins undergo a continuous exchange between various conformations on a range of timescales (femtoseconds to days) depending on the energy barrier separating the states under investigation. The complete characterisation of a protein therefore does not only require the determination of the lowest energy state but should

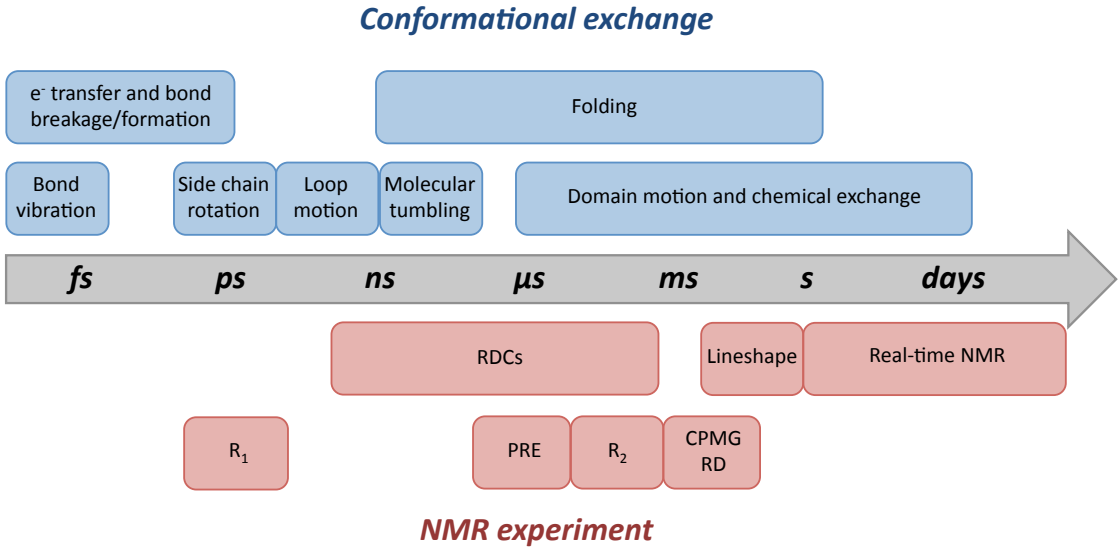

Figure 1.11: Experimental timescales accessible by NMR spectroscopy.

ideally include a structural description of each state in the conformational landscape, the population of each state and the rate and pathways of interconversion [163].

NMR spectroscopy is therefore particularly valuable to the study of protein dynamics for its high resolution provides atomistic information of local structure and dynamics, the non-perturbing nature of the NMR experiments, the large range of applications available and a solid theoretical framework permitting the interpretation of the acquired data. NMR studies have enabled detailed elucidation of chemical exchange events such as protein folding pathways [164], enzymatic cycles [165] and high energy, low population states [166]. In figure 1.11, an overview of NMR experiments and the biological exchange events they probe are summarised.

Real time NMR approaches (sometimes in combination with hydrogen-deuterium exchange (HDX)) monitor time dependent changes in resonance intensity as a probe of slow exchange events (min to days) such as *cis/trans* proline isomerisation and large scale domain movements involved in substrate binding or protein aggregation events [167, 168].

Line-shape analysis studies events that occur with an exchange rate between 10-100 ms. It relies on the monitoring of the linewidth (width at half height) of an NMR resonance at different stages of a titration (temperature, ligand etc.) to determine the kinetic properties of binding events or slow local conformational changes [169].

Carr-Purcell Meiboom-Gill (CPMG) relaxation dispersion measurements (further described in Section 2.1.6.2) allow the detailed characterisation of exchange events on a millisecond timescale. Such events include side chain reorientations, loop motions, secondary structure changes and hinged domain movements reporting on biological phenomena such

as ligand binding, folding, and enzymatic activity. CPMG relaxation dispersion experiments have been used to build a detailed picture of the DHFR catalytic cycle where the various ligand bound states transiently sample the next conformation in the enzymatic cycle confiding directionality to the enzymatic process [165]. Moreover CPMG relaxation dispersion has been used to probe low populated, so called 'invisible', states which can be both functional conformations or intermediates on the pathway to aggregation [170, 171].

Residual dipolar couplings (RDCs) provide both structural and dynamic information across a wide range of timescales. They report on the relative orientation of atoms and potential fluctuations in their orientation on a millisecond to nanosecond timescale. In stable structures, the information on relative angles between atoms within a macromolecule or between sub-units of a protein complex is extremely valuable to the determination of tertiary structure. In dynamic systems, disagreement between RDCs predicted from a known structure and the experimental values can report on details of the exchange event although deconvoluting kinetic and structural parameters can be non-trivial [172]. An exhaustive RDC dataset of ubiquitin allowed the determination of its dynamic structural ensemble showing that isolated ubiquitin populates many of its substrate-bound conformations. Binding to ubiquitin thus occurs through conformational selection [173].

Paramagnetic relaxation enhancement (PREs) measurements rely on the rapid relaxation of the NMR signal of atoms in close proximity of a free electron species to produce distance measurements. PRE measurements have proven particularly useful in the elucidation of transient interactions which escape typical structural and dynamic approaches due to their rare occurrence and short lifetimes (microseconds) and have for instance permitted a detailed characterisation of the α-synuclein unfolded structural ensemble showing that despite its intrinsically disordered nature in the absence of lipid binding partners it adopts preferred conformations driven by the charge distribution over the sequence.

Order parameters (S^2), longitudinal relaxation rate (R_1) and nuclear Overhauser effect (NOEs) measurements are reporters on the fastest timescale available to NMR (picosecond to nanosecond). This corresponds to rapid fluctuations in the local structure and typically reports on the stability of particular secondary structural elements and side chain packing.

Although the theoretical basis for the various NMR experiments has existed for decades, the application to protein dynamics has taken off in the last 20 years. Where initially specific motions on precise timescales were probed, the field has now advanced to study a protein's overall dynamic behaviour using complementary NMR and biophysical methods to report on all motions relevant to a particular biological event. For instance, the catalytic cycle of helicase motor protein SecA was shown to rely on a disorder-to-order transition in the nucleotide binding site for allosteric communication to other domains. Chemical shift perturbations demonstrated structural differences between various nucleotide bound states exist. NOE measurements showed structural elements in the nucleotide binding

site were largely disordered in the absence of nucleotide. In the presence of ADP, these structural elements were stabilised but remained dynamic as shown by differential line-broadening. The stabilisation of these elements as further characterised by isothermal titration calorimetry (ITC) resulted in structural and dynamic changes in the pre-protein binding domain testifying to communication between the two domains. The presence of ATP stabilised these structural elements fully locking the nucleotide binding site in one specific conformation [174].

1.3.2 Large macromolecular NMR studies

Our current understanding of large multi-domain proteins and protein assemblies often relies on the visualisation of components by x-ray crystallography or low resolution EM maps to report on mechanisms of activity. In particular the structural investigation of large macro-molecular assemblies is often complicated by the dynamic nature of these systems. As described above, Section 1.3.1, NMR is uniquely suited to answers this particular type of questions providing unique insights into functional mechanisms. The study of large macromolecules by conventional NMR methods is however hindered by a combination of large spectral overlap due to an increasing number of probes present in the protein and increasingly broad linewidths due to rapid relaxation of the NMR signal resulting in poorly resolved spectra [175].

At the turn of the century, considerable effort has been made towards developing new approaches in isotope labelling strategies and complementary NMR techniques. Labelling using perdeuteration approaches in combination with the selective re-introduction of protons to either NH or CH_3 groups has permitted the reduction of observable probes to a suitable level, while reducing relaxation of the NMR signal [176]. TROSY (Transverse Relaxation Optimised Spectroscopy) NMR experiments [177] exploit the specific magnetic properties of these chemical groups to skew signal relaxation in a favourable direction. The theoretical underpinnings of the TROSY effect and the technical considerations involved in selective labelling strategies in NMR studies are further described in Section 2.1.3 and Section 2.1.4 respectively. A landmark study in the development of NMR techniques suitable to the study of large protein complexes demonstrated that cross-peaks originating from GroES in a 900 kDa complex with GroEL could be readily detected using perdeuteration and TROSY-based NMR techniques [178].

The combination of perdeuteration and TROSY-based NMR experiments now permits routine structural NMR studies of systems up to 150 kDa as was demonstrated for the 82 kDa malate synthase G (MSG). More that 95% of backbone chemical shifts were assigned [179] from which secondary structure could be determined in good agreement with the crystal structure. Selective protonation of isoleucine, leucine and valine side-chain methyl

groups and their assignment [180] allowed the mapping of ligand interaction sites and their affinity through chemicals shift perturbation analysis. It was found by RDC measurements that the domain orientations of the APO state of malate synthase G resemble those of the glyoxylate-bound crystal structure while ligand-binding did not result in large scale conformational changes [181]. Further CPMG relaxation dispersion experiments found that MSG displays wide-spread millisecond dynamics [182].

1.3.3 NMR and applications in co-translational protein folding

The NMR investigation of the ribosome have shown that in large part, the 70S particle gives unobservable, broad, signals due to lack of motional averaging caused by the large size of this macro-molecular machine. The ribosome ^{15}N HSQC spectrum uniquely displays signals mapping to the highly flexible L7/L12 stalk region [183, 184] and ribosomal protein S1. It is the intrinsic flexibility of a 20 residue unstructured linker which allows the independent motion of the L7/L12 C-terminal domain relative to the ribosome (Fig 1.4).

The intrinsic motional freedom of the unstructured nascent chain relative to the ribosome mimics that of the L7/L12 ribosomal stalk region as indicated by fluorescence anisotropy measurement of the correlation times (3-7 ns) of apomyoglobin-RNCs [185]. Indeed, an extensive two-dimensional NMR study into the co-translational folding behaviour of two sequential domains of *D. discoideum* gelation factor ABP120 (ddFLN domain 5 and 6) demonstrated that the linewidths of resonances originating from the nascent chain are small enough to be detected. It was shown that the N-terminal domain 5 could reach a native conformation while domain 6, partially buried inside the ribosomal exit tunnel could not [186]. Retraction of the nascent chain into the ribosomal exit tunnel by truncation of domain 6 showed that domain 5 occupies a partially folded state with native-like properties in close proximity to the ribosome surface [65].

In similar studies, it was found that the SH3 domain of α-spectrin [187] and barnase [188] can fold to their native structure at 9, 22 and 50, and 86 residues from the PTC, respectively. The sharp linewidths detected for α-spectrin indicated that only 1% of nascent chains interact with the ribosome surface at any given time indicating the ribosome does not impose significant conformational constraints on this RNC [187]. Most recently, an elaborate study on the ddFLN5 RNC system employed a number of domain 6 linker-lengths to elucidate the rapid transition from a fully unfolded state to a native-like folded conformation with increasing distance from the PTC (Cabrita *et al.*, in preparation).

1.4 Conclusion

In this chapter, a brief description of the protein folding problem and its evolution through the past two decades has been provided. In recent years it has become apparent that the

protein folding problem cannot be disconnected from its *in vivo* conditions. The low stability of a protein's tertiary structure means that cellular factors such as crowding and interactions with other proteins can significantly affect folding states and pathways. In particular, co-translational folding constitutes a scenario that cannot be readily mimicked in isolated proteins. While the study of the molecular chaperone, trigger factor, has provide some initial insights into the role of ribosome-associated chaperones in the cotranslational folding process, a detailed understanding of the specific interaction of TF with nascent chains remains elusive.

The work presented in this thesis aims to develop a strategy for the high resolution characterisation of the dynamic interaction of TF with ribosome nascent chains complexes using NMR spectroscopy, in order to advance our mechanistic understanding of the TF chaperoning action. In first instance, the NMR behaviour of TF dimer in isolation is characterised using recent advances in selective isotopic labelling strategies and TROSY-optimised experiments and a mathematical model is developed for the interpretation of more complex equilibria. Then, using two parallel nascent chain models, the interaction of TF with isolated proteins and their equivalent nascent chains is investigated at a residue-specific level.

Chapter 2

Materials and Methods

2.1 Nuclear Magnetic Resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy has become an increasingly important tool for the high resolution structural studies of proteins. Since NMR spectroscopy probes the local magnetic environment of individual nuclei, it produces high resolution information of the chemical environment and the fluctuations therein. This can translate at a molecular level to insights ranging from structural information in the form of primary sequence connectivity, bond angles, hydrogen-bonding patterns and aromatic side chain packing to the study of local and global dynamics, low populated species and protein-protein interactions. In this Section, the NMR phenomenon is briefly described, followed by an overview of the methods by which biological NMR spectroscopy can provide detailed information about structure and dynamics over a wide range of timescales.

2.1.1 Origin of the NMR signal

At the atomic level, magnetic moments can arise from three different sources: the electron orbital angular momentum, the intrinsic spin angular momentum of the electrons and the spin angular momentum of the nucleus. In most organic molecules, having completely filled electron shells, both electron spin and orbital angular momentum sum to zero, leaving only the nuclear spin angular momentum to determine the behaviour of the atom in a magnetic field. The nuclear spin quantum number (I) of nuclei typically used in protein NMR such as ^{1}H, ^{13}C, ^{15}N and ^{35}P is 1/2. In radicals (containing an unpaired electron), the net electronic angular momentum is non-zero and therefore magnetically active, which is the basis of other spectroscopic techniques, particularly electron paramagnetic resonance (EPR, aka ESR) [189].

In the absence of an external magnetic field, the magnetic moments of the nuclei occupy random orientations resulting in a net zero magnetisation of the sample. When an external

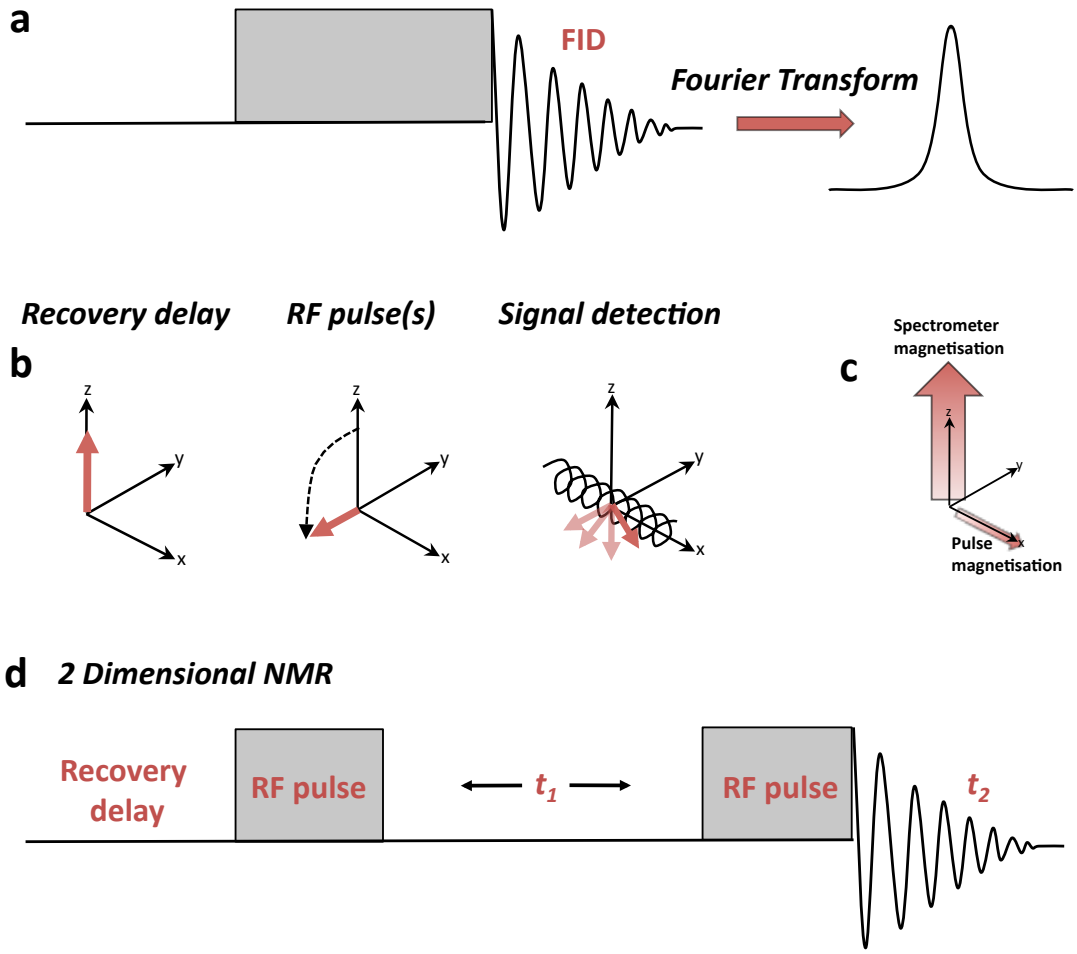

Figure 2.1: The NMR experiment. A NMR experiment follows a three step approach, schematically represented in **a**. During the delay period, equilibrium magnetisation recovers as shown in **b** (red arrow). A series of RF pulses rotate the desired magnetism into the transverse plane where it can be detected by a receiver coil. The signal is then converted by Fourier transformation to the frequency domain. **c** outlines the orientation of the magnetic fields according to the defined axis system of the spectrometer. In **d** a basic scheme for the acquisition of a two-dimensional spectrum is shown, with an additional evolution period t_1 during which the second dimension is encoded.

field, B_0, is applied, as achieved by introducing the sample into the NMR spectrometer, the distribution of nuclear spin orientations becomes biased, favouring lower energy states aligned with the magnetic field. The extent of this bias depends on the type of nucleus and the strength of the external field. For I=1/2 nuclei, the field dependence is expressed by $\Delta E = -\gamma B_0$, where γ is the gyromagnetic ratio of the nucleus. It is this bias in spin orientation that we detect as the bulk magnetic moment of the sample.

Simple NMR experiments may be described by a vector model as shown in Fig 2.1.b. Once equilibrium has been reached within the magnet (Fig 2.1.a and b), a resonant ra-

dio frequency (RF) pulse is applied to manipulate the orientation of the spins. The bulk magnetisation vector is rotated into the transverse plane (perpendicular to the external field), and the ensuing precession at the Larmor frequency, ω= -γB_0, can be detected as oscillations by the electronics of the spectrometer (Fig 2.1.a and b). The detected oscillations follow an exponential decay as the directional bias in the transverse plane dissipates (relaxes) and the spins return to their (undetectable) equilibrium distribution, aligned with the external magnetic field. The resulting acquired signal is typically converted, by Fourier transformation, to a frequency domain spectrum which is more readily interpretable. The slower the decay ('relaxation', see Section 2.1.2) of the signal, the more precisely the frequency of the oscillation can be determined and this results in a narrow frequency distribution in the spectrum. Conversely, fast relaxing signals result in broader peaks in the detected spectrum. Since the total intensity (area under a peak) depends mainly on the number of equivalent spins; a broad signal will have a lower signal-to-noise ratio than a narrow signal arising from an identical number of spins and will thus be more difficult to observe. The precise position of the peak in the frequency domain is expressed in relative terms to the frequency of DSS and is referred to as the chemical shift of the resonance [190].

In the case of two-dimensional (2D) NMR, an additional evolution period (t_1, Fig 2.1.c) is inserted after the first set of radio-frequency pulses, which encodes the magnetisation with information regarding the second frequency dimension, followed by a mixing period where RF pulses transform the magnetisation into an observable state (Fig 2.1.c). A single 2D experiment corresponds to a series of one dimensional (1D) experiments with increasing evolution periods (t_1). The two dimensions of a 2D experiment are each associated with one of the two time variables: the evolution period (t_1) and the detection period (t_2). Both are then converted to the frequency domain through Fourier transformation. The acquisition of a 2D NMR spectrum is more time consuming than 1D NMR since it requires the repeated acquisition of a number of 1D spectra, each with increasing evolution periods. Important trade-offs therefore have to be made in the experimental design. In particular, the maximum evolution time (t_1) in the indirect dimension can severely affect the resolution of the acquired NMR spectrum, since longer times are needed to discriminate between oscillations at similar frequencies yet longer times produce less signal due to relaxation. Longer t_1 times also directly affect the duration of the experiment. Further considerations concern the total frequency range monitored ('spectral width'). The larger the spectral width, the closer together time-points in the indirect dimension are required to be, increasing the number of total 1D spectra that need to be acquired and thus increasing experimental time. The spectral width is therefore typically set to the smallest range possible in order to either improve resolution or reduce experiment time. For samples of low signal-to-noise, resolution (in the form of acquisition time) is often

traded for additional repeats of the same experiments which can be summed to improve signal intensities.

Within a molecule, magnetic nuclei can interact by means of the electrons in covalent bonds linking them. This phenomenon, J-coupling, results in splitting of the peaks observed in the spectrum for both nuclei. In protein NMR, this is typically undesirable since it adds to the crowding of the spectrum and thus the splitting is often collapsed using decoupling pulses during acquisition of the signal, resulting in a single average peak in final spectrum (See Fig 2.3).

2.1.2 Relaxation

The exponential decay of the observed NMR signal results from the relaxation of coherences towards their equilibrium state in the spectrometer, aligned with the B_0 field. A distinction is made between transverse and longitudinal relaxation. Longitudinal relaxation, quantified by the rate constant R_1, corresponds to the nonadiabatic realignment of spins with the external magnetic field (Fig 2.2.a). Transverse relaxation, described by the rate constant R_2, reflects the dephasing of the magnetisation in the transverse plane due to fluctuations in the frequencies of individual spins (Fig 2.2.b).

As relaxation results in net zero transverse magnetisation, and thus an undetectable state, the faster transverse relaxation rate (Fig 2.2.c) limits the time during which the NMR signal can be acquired. The slower longitudinal relaxation rate imposes a minimal delay period between experiment during which the longitudinal magnetisation recovers but no transverse magnetisation is detected. The rate of both relaxation mechanisms depends on the global rotational diffusion of the molecule, characterised by the rotational correlation time τ_c (Fig 2.2.c) of the molecule under investigation and chemical exchange processes. The rotational correlation time and molecular weight can be estimated from the hydrodynamic radius (r_h), assuming a spherical particle, using the Stokes-Einstein-Debye relation:

$$\tau_c = \frac{4\pi\eta r_h^3}{3k_B T} \tag{2.1.1}$$

where k_B is the Bolzmann constant, T is the temperature and η is the viscosity of the sample and

$$MW = \frac{4\pi N_A \left(r_h - r_w\right)^3}{3\nu_0} \tag{2.1.2}$$

where N_A is Avogadro's number, r_w is the thickness of the hydration shell (estimated at 0.2 nm [191] and ν_0 is the specific volume of the particle (estimated at 7.3×10^{-7} $\mathrm{m}^3.\mathrm{g}^{-1}$ for proteins and 5.5×10^{-7} $\mathrm{m}^3.\mathrm{g}^{-1}$ for nucleic acids) [192].

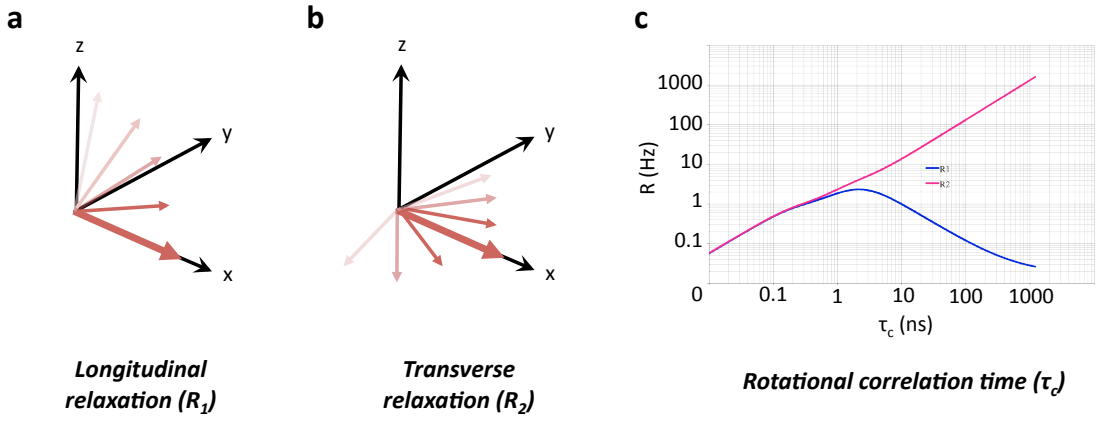

Figure 2.2: Relaxation (**a**) Longitudinal relaxation corresponds to the recovery of the alignment with the external magnetic field. (**b**) Transverse relaxation reflects the dephasing of coherences in the transverse plane. (**c**) Heteronuclear relaxation rates (R_1 and R_2) as a function of correlation time (τ_c) calculated for a ^{1}H spin at 700MHz. Adapted from Helene Launay, doctoral thesis, 2011.

The rate of transverse relaxation (R_2) has a significant impact on the physical appearance of the NMR spectrum. The linewidth of NMR signals (the width at half height) in Hz corresponds to R_2/π.

2.1.3 Transverse Relaxation-Optimised Spectroscopy

Since relaxation rates are strongly dependent on the size of the studied systems, the NMR investigation of large macromolecules (>25 kDa) requires overcoming rapid transverse relaxation rates that lead to broad NMR signals. The exploitation of differential relaxation behaviour, the TROSY effect (Transverse Relaxation-Optimised Spectroscopy) [177], for certain chemical groups such as backbone amides and side-chain methyl groups has permitted significant improvements in spectral quality and sensitivity, in the study of such larger systems. The TROSY effect originates from the interference (or cross-correlation) of multiple sources of relaxation which can have an additive or destructive effect on the overall strength of the local magnetic field, based on the orientation of a coupled spin. This results in a stark difference in the linewidths of the peaks in the J-coupling multiplet of the observed nucleus, here illustrated for amide groups, where the fast relaxing orientation results in a broad peak and the slow relaxing state produces a narrow peak (2.3.a). Ideally, this results in the broadening beyond detection of three of the multiplet peaks and a maximised height for the last peak corresponding to the slow relaxing species only (Fig 2.3.a). The multiplet is not collapsed which results in a net loss of 75% of the signal (area under the peak) but the gain in signal/noise ratio (20-50 fold [193]) associated with the reduction in linewidth of one of the peaks largely compensates for this loss in proteins

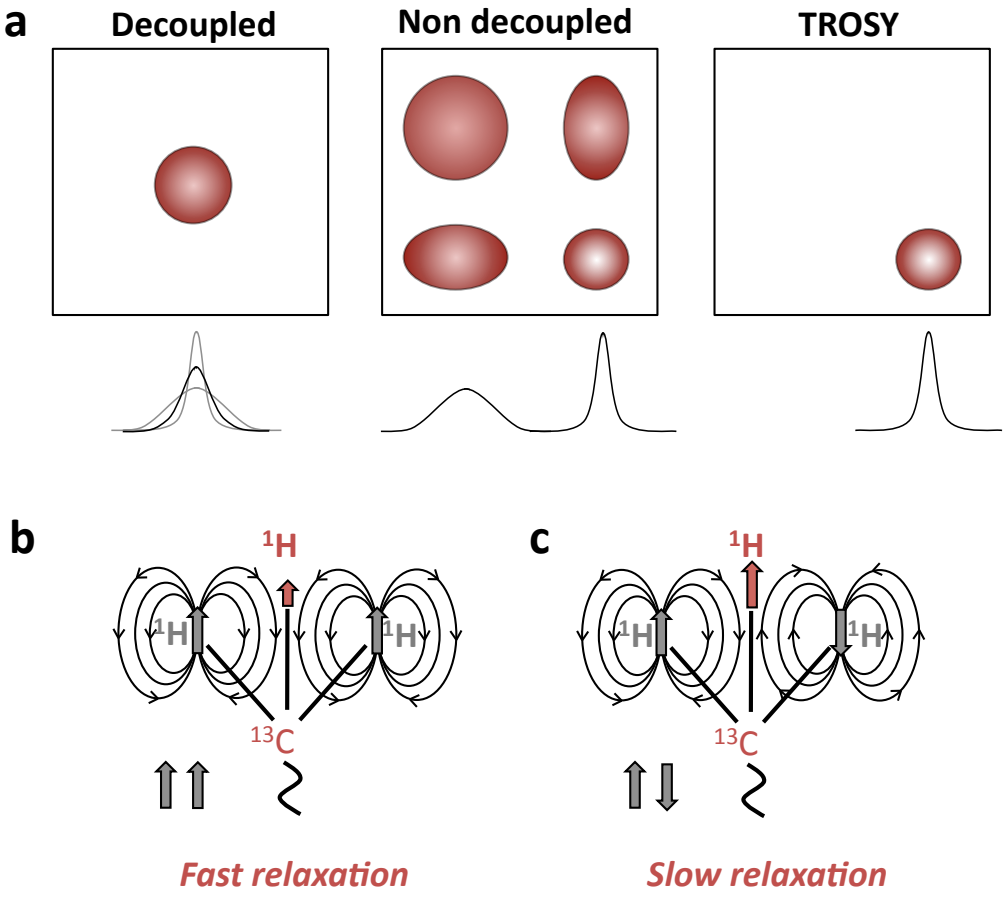

Figure 2.3: Transverse Relaxation Optimised Spectroscopy (TROSY). Typical 2D NMR ^{1}H-^{1}5N HSQC and HMQC experiments apply decoupling pulses during acquisition collapsing the signal to a resonance of a linewidth, the average of all components (**a**). Without decoupling, the resonance of an ^{15}NH spin pair is split into four components which each relax at a different rate resulting in differential linewidths. The TROSY effect skews this imbalance in linewidths to a maximum where only the slowest relaxing state remains visible thus avoiding the averaging of linewidths that happens during traditional NMR experiments. (**b**) Fast relaxation of methyl protons is induced by the parallel alignment of neighbouring spins. (**c**) Neighbouring spins of opposing signs do not induce relaxation at the observed methyl proton.

greater than 15-20 kDa [175].

2.1.4 Selective isotopic labelling strategies

Early approaches to making proteins accessible to study by NMR spectroscopy involved enrichment in otherwise low natural abundance magnetic nuclei such as ^{13}C and ^{15}N, by the heterologous expression in isotopically-enriched media, enabling heteronuclear 2D and 3D experiments by resolving spectral overlap. Such isotope enrichment strategies have permitted the detailed investigation of systems up to 25kDa [194, 195].

However, as described (Section 2.1.2), the increasing size of the systems studied is correlated with increasing rotational correlational times and leads to rapid transverse relaxation. For larger macromolecules, both the spectral overlap and fast relaxation become increasingly problematic. Fractional deuteration or perdeuteration of the sample is a first step towards overcoming both of these problems since deuterium is a silent nucleus in most NMR experiments and effectively suppresses a large number of signals. Moreover, the relaxation rates associated with deuterium are significantly slower than for protons due to the smaller gyromagnetic ratio ($\gamma_D/\gamma_H = 0.15$) [196] resulting in narrower linewidths of remaining resonances [197].

Recently developed powerful strategies for the investigation of large macromolecular assemblies rely on the selective re-introduction of protons into the protein by means of back-exchange of accessible groups such as backbone amides [198, 199] or selective labelling of amino acid side chains using isotopically labeled metabolic precursors during expression[177] described further in Section 3.3.

2.1.5 Methyl-TROSY NMR of large macromolecules

Over the past decade, the theoretical limit on the molecular weight of systems within the reach of high resolution NMR study has been extended, using a combined strategy of selective isotopic labelling and TROSY-enhanced signal detection. The Kay lab has pioneered many of these developments [196, 200].

The essence of this strategy is to prepare fully perdeuterated protein samples in order to suppress all proton-based sources of relaxation. Amino acid side-chain methyl groups are then selectively ^{1}H-^{13}C labeled making these the only observable species within the protein. The choice of methyl groups is justified by 1) the three-fold degeneracy of the methyl proton tripling the signal for this particular resonance, 2) the long(er) side-chains of methyl-group containing amino acids, such as Ile, Val and Leu, conveying additional flexibility compared to the general correlation time of the protein, resulting in slower relaxation of the signal[201, 202], 3) these hydrophobic groups are typically buried within the protein core, making them excellent probes of structural features and integrity [203, 204] and 4) the methyl TROSY effect.

The TROSY effect of a methyl group proton is caused by the spin orientations of the two neighbouring protons: when both are aligned to the external magnetic field in the same direction, their magnetic fields enhance the magnetic field of the observed proton resulting in rapid relaxation, while the magnetic fields of two opposing spins cancel out, resulting in slow relaxation (Fig 2.3.b and c). In particular in large macromolecules, the difference between the fast and slow relaxing components is considerable, leading to the broadening beyond detection of resonances affected by the fast relaxation mechanism while

slow relaxing signals are enhanced.

Adaptation of conventional pulse sequences to exploit the TROSY effect by preventing the mixing of slow and fast relaxing species during the experiment has enabled the high resolution structural investigations of large molecular assemblies such as the 20S proteosome [205, 206] as described in Section 1.3.2.

2.1.6 NMR structural and dynamics methods

NMR spectroscopy is particularly valuable in the high resolution study of dynamics systems as was discussed in Section 1.3.1. Here, the theoretical basis of specific structural and dynamic NMR experiments used throughout this study are detailed.

2.1.6.1 Chemical exchange and the NMR timescale

The appearance of NMR spectra are highly sensitive to chemical exchange - the dynamic interconversion of an NMR nucleus between at least two distinct chemical environments such as two conformational states, A and B. Due to the differences in chemical environment, each state may have distinct resonance frequencies Ω_A and Ω_B. The appearance of the NMR spectrum will be dependent on the rate of exchange between the two states, k_{ex}=k_{AB}+k_{BA}, the extent to which each state is populated, P_A and P_B and the frequency difference, $\Delta\Omega$=$|\Omega_A - \Omega_B|$ (related to the chemical shift difference, $\Delta\Omega$= $2\pi\nu_0\Delta\delta/10^6$ where ν_0 is the Larmor frequency, $\Delta\delta$ is the chemical shift difference in ppm) as shown in Fig 2.4 [163].

Exchange processes fall into three distinct regimes. Under slow exchange ($k_{ex} \ll \Delta\Omega$), distinct signals are observed for both states A and B at their respective chemical shifts, Ω_A and Ω_B. The intensities of the signals are directly proportional to the populations of A and B because no significant interconversion between the two states occurs on the NMR timescale. Conversely, in the fast exchange limit ($k_{ex} \gg \Delta\Omega$), the rapid interconversion of states leads to the averaging of the NMR signals originating from state A and B, resulting in a peak at a population-weighted average chemical shift position with a total intensity, the sum of A and B. In between the slow and fast exchange limits, in the intermediate exchange regime ($k_{ex} \approx \Delta\Omega$), the observed spectrum is severely affected by exchange broadening. The extent of broadening is strongly dependent on the chemical shift difference and on the kinetics of the exchange process and forms the basis of lineshape analysis and CPMG relaxation dispersion methods (see Section 2.1.6.2). In the fast-intermediate regime (below the coalescence point where k_{ex} =$\Delta\Omega$), the exchange contribution to the relaxation rate corresponds to R_{ex}=$\frac{P_A P_B \Delta\Omega^2}{k_{ex}}$ while in the slow-intermediate regime (above coalescence), the exchange contribution to the relaxation rate depends on the off-rate only: $R_{ex}^A = P_B k_{ex}$.

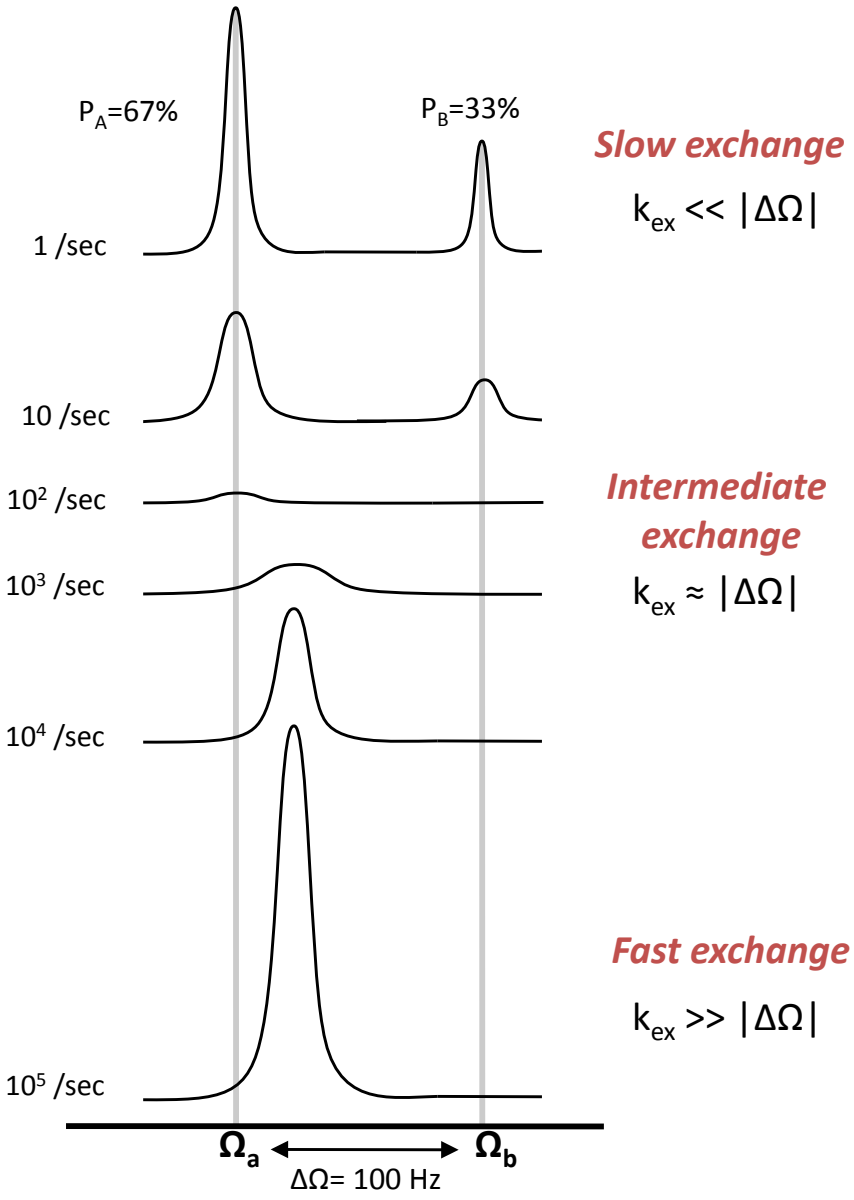

Figure 2.4: Chemical exchange and NMR timescales. The physical appearance of the NMR spectrum is highly sensitive to chemical exchange as illustrated here by the example of two-state exchange. At slow timescales ($k_{ex} \ll \Delta\Omega$), signals for both states are observed of distinct linewidth and intensity at their respective chemical shifts. Under fast exchange conditions ($k_{\mathrm{ex}} \gg \Delta\Omega$) a single resonance is observed reflecting the population weighted average chemical shift, intensity and linewidth. On intermediate timescales, the linewidth of observed resonances is severely enhanced by chemical exchange. In the slow-intermediate regime, the linewidth of the resonance originating from the minor population B is enhanced by the rate of interconversion to the major state A (k_{BA}) while in pure intermediate exchange, the linewidth of the observed resonances is affected by both k_{AB}, k_{BA} and the population distribution. Adapted from [207]

Since the exchange regime is defined as a function of both the exchange rate k_{ex} and the frequency difference $\Delta\Omega$, it is possible to alter the relative timescale of the exchange by altering the magnetic field strength B_0. Since $\Delta\Omega$ scales linearly with B_0, an increase in B_0 will decrease the relative timescale of the exchange (for k_{ex} is unaffected by the applied magnetic field) [163].

2.1.6.2 Carr-Purcell Meiboom-Gill Relaxation dispersion

The exchange broadening detected in the intermediate exchange regime contains valuable information on the millisecond-second timescale dynamics of the protein. Carr-Purcell Meiboom-Gill relaxation dispersion (CPMG RD) NMR exploits this contribution of the exchange to the detected relaxation rates to extract kinetic, thermodynamic and structural information about the underlying biological event [208, 209].

CPMG RD experiments manipulate the exchange contribution to the transverse relaxation (R_{ex}) by applying a variable number of refocussing pulses during a fixed relaxation period (T_{CPMG}). These refocussing pulses or 'spin echo' elements are typical features of NMR pulse sequences that refocus magnetisation by means of symmetrical delay periods (τ) around a 180° pulse in the transverse plane. In the absence of exchange, magnetisation precessing in the transverse plane will return to its starting position at the end of the spin echo period resulting in a sharp, focussed peak (see Fig 2.5.b (grey line)) in the NMR spectrum. However, if exchange between two states of the NMR probe, each associated with their own precession frequency (chemical shifts Ω_{A} and Ω_{B}), occurs during the spin echo, the magnetisation is not refocussed, resulting in a broad peak (Fig 2.5.a and b). Increasing the number of spin echo elements (by reducing the symmetrical waiting periods) reduces the likelihood that exchange occurs before the magnetisation is refocussed thus resulting in a sharper peak. The degree of refocussing as a function of the CPMG frequency contains information about the exchange rate (k_{ex}), the population distribution (p_{A}+p_{B}=1) and the chemical shift difference ($\Delta\Omega$) between the two states, A and B.

In practice, peak intensities are measured at various CPMG frequencies, converted to transverse relaxation rates (R_2^{obs}) by comparison with peak intensities in the absence of CPMG period (T_{CPMG}) and fit to a model for two-state equilibrium described by the Carver-Richards equations which express the observed relaxation rates (R_2^{obs}) in terms of the CPMG frequency (ν_{CPMG}=$1/4\tau$, where 2τ is the delay between two spin echo pulses), the exchange rate (k_{ex}), the population distribution (p_{A} and p_{B}), the frequency difference ($\Delta\Omega$=$\pi \times \Delta\omega$) and the relaxation rate in absence of exchange (R_2^0) [210]:

$$R_2^{obs} = \frac{1}{2}\left(R_{2A}^0 + R_{2B}^0 + k_{ex} - \left(\frac{1}{2\tau}\right)\cosh^{-1}\left(D_+\cosh\left(\eta_+\right) - D_-\cos\left(\eta_-\right)\right)\right) \quad (2.1.3)$$

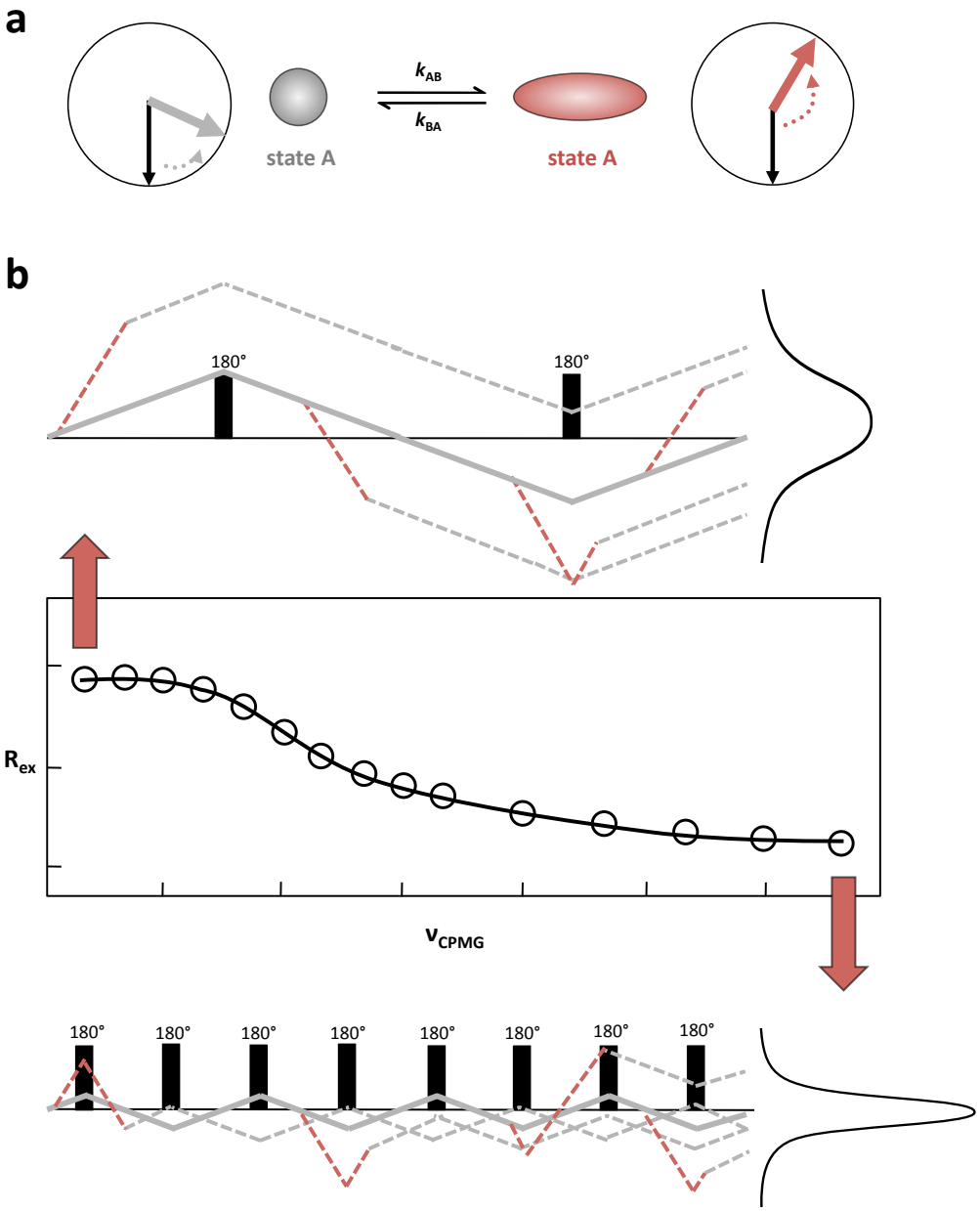

Figure 2.5: Carr-Purcell Meiboom-Gill (CPMG) Relaxation dispersion. (**a**) In the case of two state exchange between state A and state B each state has its own associated procession frequency (chemical shift) due to the differences in chemical environment. (**b**) The CPMG RD experiment uses spin echo pulses (black rectangles) to refocus magnetisation to its starting position. In the absence of exchange (grey) this refocussing is complete. In the presence of exchange (dotted lines), the magnetisation is not refocussed. Increasing the frequency of the spin echo pulses to exceed the exchange rate improves the extent of refocussing (bottom). The extent of refocussing as a function of the spin echo frequency (ν_{CPMG}) can be fitted to the Carver-Richards equations or numerical models (see text) to extract the chemical shift differences and populations of the two states as well as the exchange rates.

$$D_{\pm} = \frac{1}{2}\left(\pm 1 + \frac{\Psi + 2\Delta\omega^2}{\sqrt{\Psi^2 + \zeta^2}}\right) \tag{2.1.4}$$

$$\eta_{\pm} = \left(\frac{2\tau}{\sqrt{2}}\right)\left(\pm\Psi + \left(\Psi^2 + \zeta^2\right)^{1/2}\right)^{1/2} \tag{2.1.5}$$

$$\Psi = \left(R_{2A}^0 - R_{2AB}^0 - p_A k_{ex} + p_B k_{ex}\right)^2 - \Delta\omega^2 + 4 p_A p_B k_{ex}^2 \tag{2.1.6}$$

$$\zeta = 2\Delta\omega\left(R_{2A}^0 - R_{2AB}^0 - p_A k_{ex} + p_B k_{ex}\right) \tag{2.1.7}$$

Carver-Richards analysis permits the extraction of structural information from the CPMG RD data in the form of the chemical shifts of state A and B. Often, comparison with NMR spectra of known conformational states such as ligand-bound forms permits the identification of the minor conformation. Kinetic and thermodynamic information can be derived from the determined exchange rate and population distribution. In particular, the temperature dependent behaviour of the kinetics as determined by CPMG RD experiments can report on the activation energy of the exchange process by means of Arrhenius analysis. Finally the fitted relaxation rate in the absence of exchange, R_2^0, reports on ps-ns timescale motions and can be used to determine the exchange broadening contribution to the relaxation R_2^{ex}=R_2^0-R_2^{obs}. R_2^{ex} values are useful as semi-quantitative descriptions of μs-ms dynamics in first instance since they report on site-specific events without assuming a model for the molecular motion they report on.

Under fast exchange conditions ($k_{\mathrm{ex}} \gg \Delta\Omega$) the Carver-Richards equations can be simplified to :

$$R_2^{obs} = R_2^0 + \left(\frac{\overbrace{p_A p_B \Delta\omega^2}^{\Phi_{ex}}}{k_{ex}}\right)\left(1 - \frac{2\tanh\left(k_{ex}\tau\right)}{2k_{ex}\tau}\right) \tag{2.1.8}$$

where the population distributions, p_A and p_B and the frequency difference $\Delta\omega$ appear only as a single parameter Φ_{ex} and can thus not be determined individually. In the fast exchange regime, CPMG experiments are thus limited to reporting on the exchange rate, k_{ex} of the chemical process only.

The convolution of parameters under fast exchange conditions and the fact that in practice often many parameter sets fit the CPMG dispersion curves (containing experimental uncertainties) equally well, severely hinders the analysis of CPMG RD data. Accurate analysis of CPMG RD data therefore often requires a number of independent data sets at different magnetic field strengths. The magnetic field strength uniquely affects the chemical shift difference, leaving kinetic and thermodynamic parameters unaffected. This reduces the number of independent parameters, permits the determination of the

chemical shift difference independently from the kinetic parameters and reports on the exchange regime of the resonance by means of the parameter α which varies between 0 (slow exchange) to 2 (fast exchange). This is because, in the fast exchange regime (equation 1.1.6), the chemical shift difference is squared while in intermediate exchange it scales linearly (equations 1.1.1-1.1.5). In addition to variation of the field strength, temperature can be used to modulate the kinetics and thermodynamics of the exchange process while the chemical shift difference is assumed to remain approximately constant. This assists towards deconvoluting the kinetic parameters from the chemical shift difference and can alter the exchange regime, preferentially away from the fast exchange regime. Finally, the CPMG RD curves of multiple resonances in the NMR spectrum can be fitted simultaneously, assuming all the observed exchange broadening originates from the same chemical process, further reducing the number of fitting parameters.

2.1.6.3 Nuclear spin relaxation

The return to equilibrium of the magnetisation after perturbation by RF pulses during the NMR experiment is governed in large part by random molecular motions on the ps-ns timescale (see Section 2.1.2 for further introduction to relaxation). Biological phenomena that occur on this timescale include bond-vibrations, side-chain re-orientations, random coil and loop motions as well as rotational diffusion of the whole molecular with correlation time τ_c. Motions on a ps-ns timescale are typically not described as the interconversion between two states but are quantified by generalised order parameters (S^2). The order parameter S^2 ranges in value from 0 (disordered) to 1 (rigid) and is often considered to reflect the amplitude of motion within a cone, formed by the bond vector, such that increasing rigidity corresponds to a smaller cone angle θ. The motions of the side chain methyl groups are highly restricted by side chain packing in the hydrophobic core of proteins and thus directly report on the rigidity of the protein fold at that position. Typically S^2 values for methyl groups in folded proteins are above 0.6 [211].

Since various sources of fluctuation in the local magnetic field affect both transverse and longitudinal relaxation rates, often a number of measurements are required to extract the effects of one particular motion. In the case of backbone NH flexibility on a ps-ns timescale, typically R_1, R_2 and heteronuclear NOE measurements are made and combined to extract S^2 parameters and a description of the rotational diffusion in the form of the correlation time (τ_c).

In the case of side chain methyl groups as used throughout this study, S^2_{axis}, which describes the amplitude of motion of the methyl three-fold symmetry axis, can be calculated

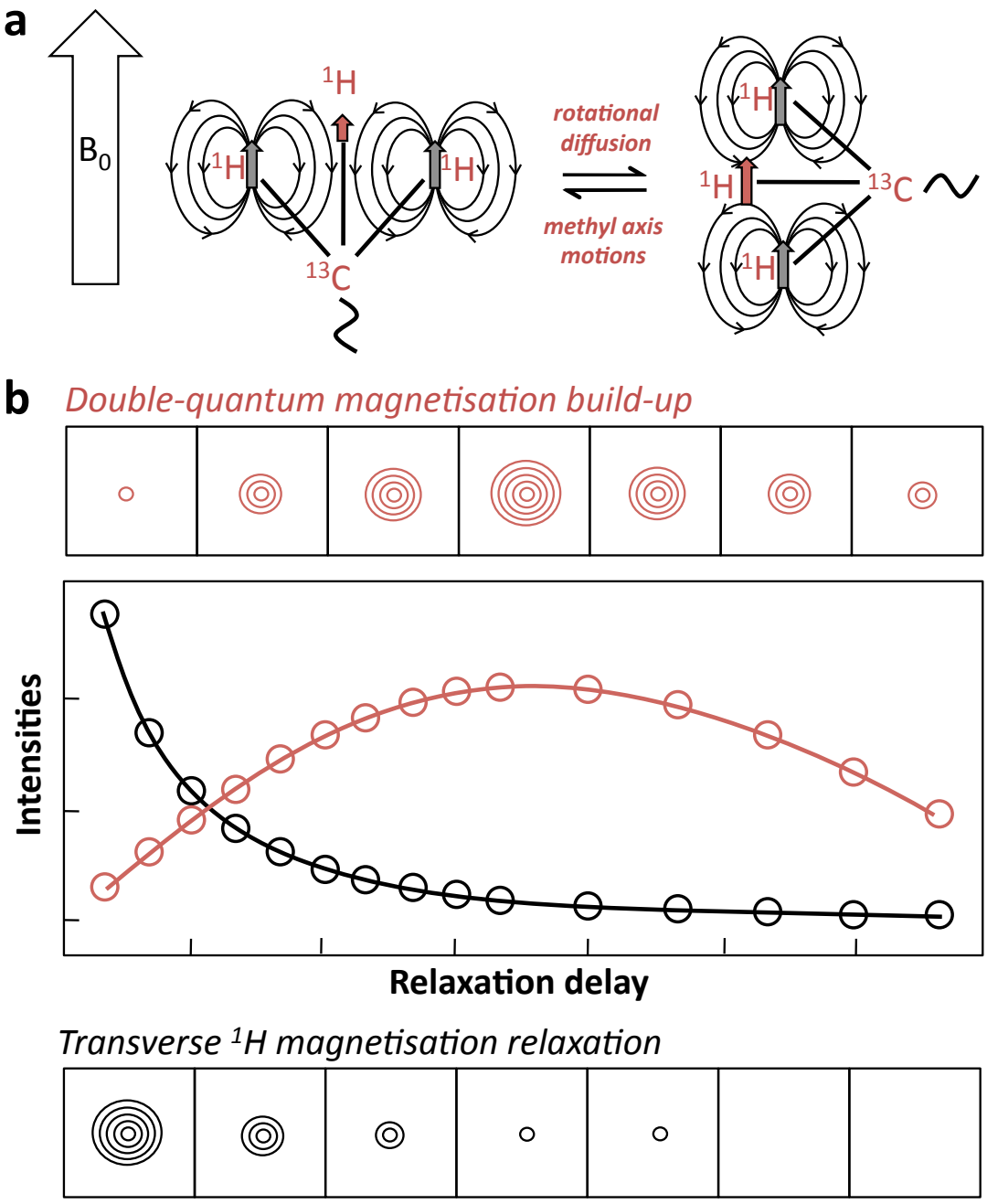

Figure 2.6: Intra-methyl ^{1}H-^{1}H dipolar cross-correlated relaxation. (**a**) As the methyl axis reorientates relative to the external magnetic field B_0, neighbouring protons are influenced differently by the magnetic field of a proton nucleus. While in a parallel conformation, the local magnetic field at a neighbouring proton is reduced, in a perpendicular orientation the local field is enhanced. It is the fluctuations thus induced in the local magnetic field that causes relaxation through what is called intra-methyl ^{1}H-^{1}H dipolar cross-correlated relaxation. (**b**) The rate of intra-methyl ^{1}H-^{1}H dipolar cross-correlated relaxation can be determined from the ratio of double-quantum magnetisation build-up and the transverse relaxation rate as measured by specific NMR experiments described in [211]. Adapted from [212].

from the intra-methyl ^{1}H-^{1}H dipolar cross-correlated relaxation rate (η) [213]:

$$\eta = \frac{R^F_{2,H} - R^S_{2,H}}{2} \approx \frac{9}{10} \left[P_2 \left(\cos \theta_{axis.HH}\right)\right]^2 \frac{S^2_{axis} \gamma^4_H \hbar^2 \tau_c}{r^6_{HH}} \tag{2.1.9}$$

where γ_{H} is the proton gyromagnetic ratio, r_{HH} is the distance between pairs of methyl protons (1.813A), $P_2(x)$=(1/2)(3x^2-1), $\hbar$ is the reduced Planck constant and $\theta_{\mathrm{axis,HH}}(90^\circ)$ is the angle between the methyl three-fold symmetry axis and a vector connecting a pair of the methyl proton nuclei.

The intra-methyl dipolar cross-correlated relaxation rate (η) describes the effect of the magnetic moments of the neighbouring protons in the methyl group on the observed proton as shown in Fig 2.6. In Fig 2.6.a, all methyl proton magnetisation is aligned with the magnetic field. The magnetisation of the two neighbouring protons slightly enhance that of the observed proton when the methyl axis is parallel to the external magnetic field B_0 as shown by the red arrows relative to the grey arrows which represent the magnetisation in the absence of neighbouring atoms. When the axis re-orientates perpendicular to the magnetic field, through rotational diffusion (τ_c) or local motions (reported by S^2), the neighbouring proton magnetisation now cancel out and no effect is detected by the observed proton. The motions in the methyl-axis orientation thus result in fluctuations in the magnetic field of the observed methyl-proton leading to the dephasing of the coherence ie transverse relaxation.

In practice, the intra-methyl dipolar cross-correlated relaxation rate can be measured with the greatest accuracy from the ratio of the build-up of double-quantum coherences (I_a) and the transverse relaxation rate (I_b) using the following equation [211]:

$$\frac{I_a}{I_b} = \frac{-0.5\eta \tanh\left(\sqrt{\eta^2 + \delta^2} T\right)}{\sqrt{\eta^2 + \delta^2} - \delta \tanh\left(\sqrt{\eta^2 + \delta^2} T\right)} \tag{2.1.10}$$

where η is the intra-methyl ^{1}H-^{1}H dipolar cross-correlated relaxation rate and δ is a cross-relaxation parameter that depends on the proximity of external protons to the methyls group in question. The fitted value of η can be substituted in to equation 2.1.9 to give S^2_{axis}.

2.1.6.4 Cross-saturation by spin diffusion

Cross-saturation experiments are a structural NMR method that permits the semi-quantitative mapping of complex interfaces, particularly when the information contained in chemical shift perturbations is complicated by further structural changes upon complex formation away from the interface such as domain hinge motions etc. Cross-saturation experiments

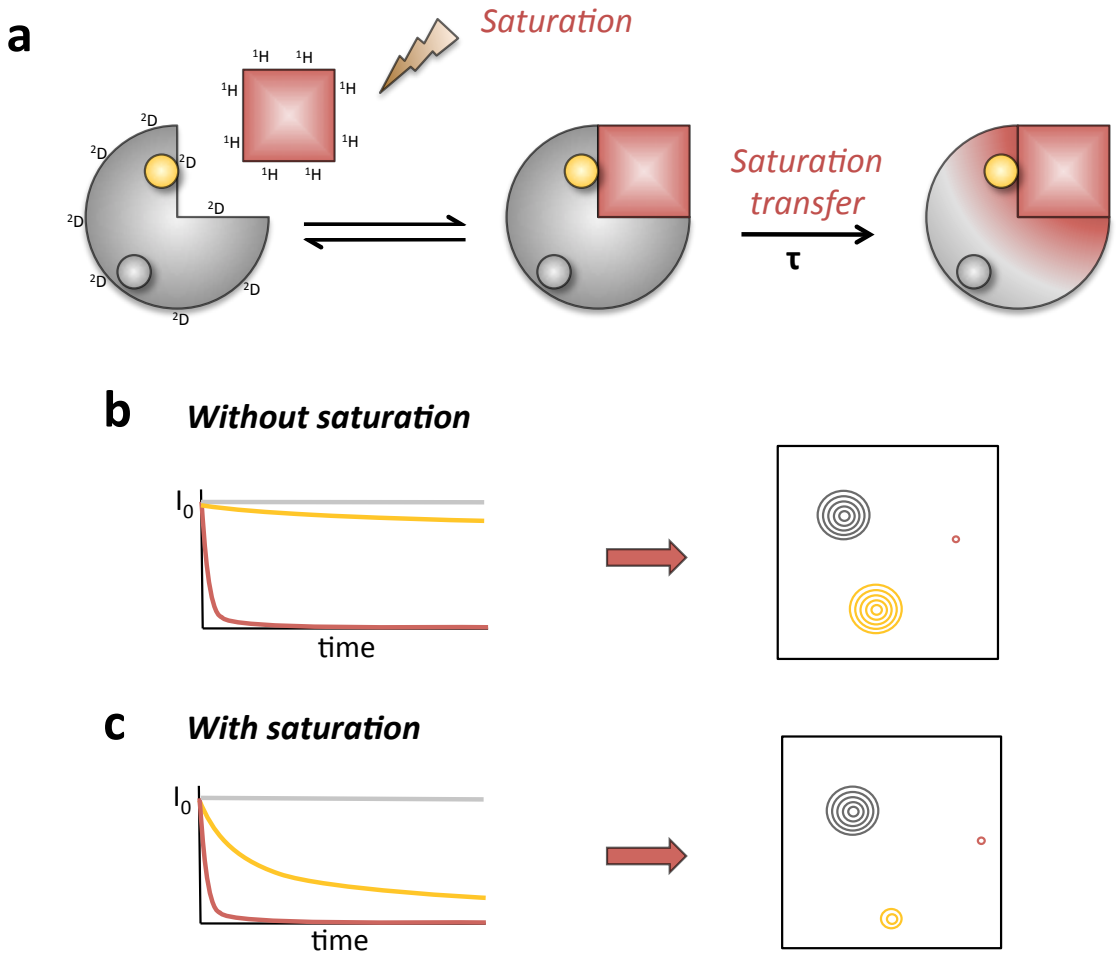

Figure 2.7: Cross-saturation. (**a**) Cross-saturation experiments exploit the differential relaxation properties of protons (^{1}H, red) and deuterons (^{2}D, grey) in a magnetic field to map protein interaction surfaces. While the magnetisation originating from the protonated subunit is saturated, the magnetisation of the deuterated partner is unaffected. During the recovery delay period, the saturation can be transferred across the binding interface through spin diffusion. (**b**) Exponential decay and two dimensional resonances of spins in different parts of the protein complex in the absence of saturation pulse. (**c**) Exponential decay and two dimensional resonances of spins in different parts of the protein complex in the presence of saturation pulse.

are based on the selective saturation of magnetisation of one subunit in the complex. Saturation refers to a state where magnetisation has been equally distributed over the high (down) and low (up) energy spin states in the bulk magnetic field. The bias in the magnetisation orientation that is typically observed in an NMR experiment is thus entirely lost. This can be achieved by applying low power selective RF pulses during the recycle period (Fig 3.19) [214, 215].

Cross-relaxation leads to the transfer of saturation of protons across the binding in-

terface which is monitored as intensity changes of the NMR signals of the binding partner (Fig 3.19). Cross-saturation is a short range effect with r^{-6} dependence reliant on the dipole-dipole interaction of nearby spins. Spin diffusion can increase the distance of the saturation transfer and interferes with quantitative analysis of cross-saturation data. Typically, the saturated subunit is unlabelled and protonated while the observed sub-unit is deuterated and selectively labelled according to one of the schemes described Section 2.1.4. This ensures that the saturation of protons is selective and that the cross-relaxation phenomenon is limited to the complex interface [216, 217].

2.1.6.5 Paramagnetic relaxation enhancement

Paramagnetic relaxation enhancement (PRE) experiments as used throughout this study are a useful tool for the determination of medium range distances within proteins and protein-complexes. PRE measurements rely on the selective introduction of a paramagnetic centre in the form of paramagnetic transition metal or nitroxide radical. While in metal binding proteins the native binding site can be exploited, the study of typical proteins requires the chemical modification of cysteine residues with lanthanide-binding tags or nitroxide spin labels as sources of free electron spins.

Unpaired electrons can cause large enhancements of relaxation rates in nearby nuclei due to the r^{-6} dependence of the PRE effect. Mediated by dipole-dipole interactions, this effect is 660 fold stronger than the effect of neighbouring protons due to the much larger electron gyromagnetic ratio. The paramagnetic relaxation enhancement of transverse relaxation (Γ_2) is defined as [217] :

$$\Gamma_2 = R_2^{\text{para}} - R_2^{\text{dia}} \tag{2.1.11}$$

where R_2^{para} and R_2^{dia} are the transverse relaxation rates of the paramagnetic and diamagnetic states of the macromolecule respectively. The diamagnetic relaxation rates are measured from the reduced state of the nitroxide label or in the absence of metal ions. This substraction eliminates additional contributions to the relaxation rates that are common to both states.

Experimental relaxation rate measurements are time-consuming since sufficient time-points need to be sampled for accurate analysis. The measurement of both R_2^{para} and R_2^{dia} are however often derived from measurements at just two time-points. Assuming exponential decay between these time points, the PRE rates (Γ_2) can be derived from the measured intensities values [218],

$$\Gamma_2 = \frac{1}{T_{\text{b}} - T_{\text{a}}} \ln \frac{I_{\text{dia}}(T_{\text{b}})\, I_{\text{para}}(T_{\text{a}})}{I_{\text{dia}}(T_{\text{a}})\, I_{\text{para}}(T_{\text{b}})} \tag{2.1.12}$$

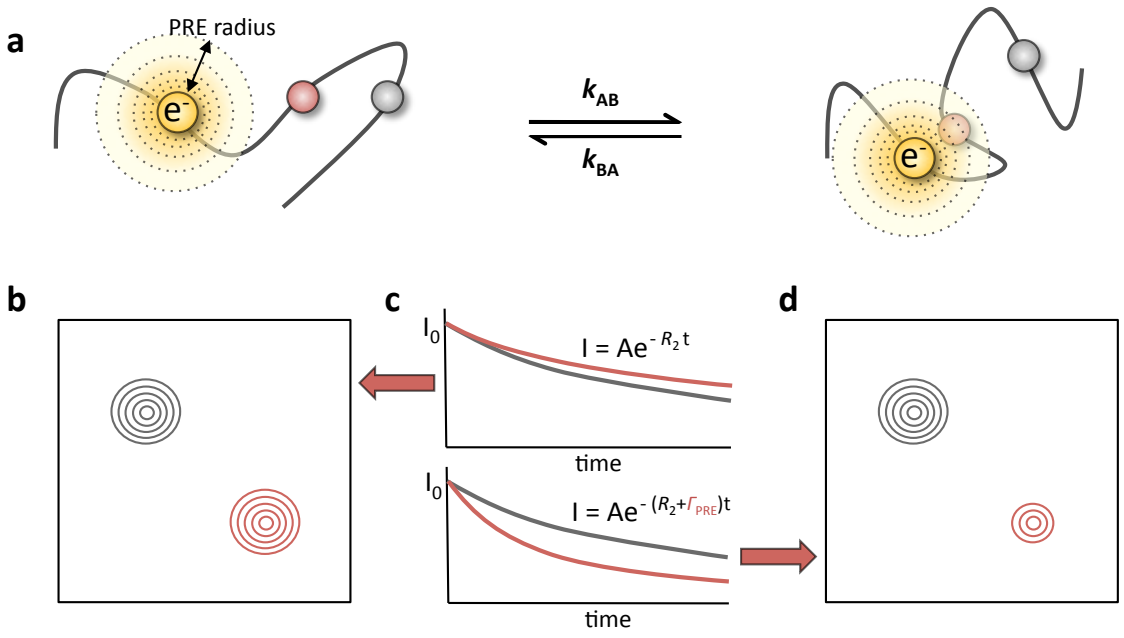

Figure 2.8: Paramagnetic relaxation enhancement. Free electron species severely enhance the relaxation rates of nearby nuclei. The r^{-6} dependence of this effect makes it a valuable tool in distance determination and to characterise transiently populated species as shown in **a**. (**b** and **d**) The effect of the PRE on the appearance of a 2D NMR spectrum.(**c**) PRE effect of free electron on the relaxation rates of a nearby (red) and removed (grey) nucleus.

where I_{para} and I_{dia} are the measured intensity values for the paramagnetic and diamagnetic samples and T_{a} and T_{b} the two time points of the relaxation experiment.

The PRE relaxation rates can be converted to distance restraints using the equation below derived from the Solomon-Bloembergen equations [219]:

$$r = \left[\frac{K}{\Gamma_2}\left(4\tau_c + \frac{3\tau_c}{1+\omega_{\text{H}}^2\tau_c^2}\right)\right]^{1/6} \tag{2.1.13}$$

with

$$K = \frac{1}{15} S\left(S+1\right)\gamma^2 g^2 \beta^2 \tag{2.1.14}$$

where r is the distance between the electron and the nuclear spin monitored, ω_{H} the Larmor frequency of the proton (the frequency at which the proton precesses in the magnetic field), S is the electron spin quantum number, γ is the gyromagnetic ratio of protons, g is the electronic g-factor and β is the Bohr magneton.

The effective correlation time τ_c of the electron-nucleus vector is defined as [217] :

$$\frac{1}{\tau_c} = \frac{1}{\tau_r} + \frac{1}{\tau_e} \tag{2.1.15}$$

where τ_r is the rotational correlation time of the protein and τ_e is the electron relaxation time [220]. The strength of the PRE effect is thus dependent on the electron relaxation rates and therefore varies based on the choice of free radical. While distances up to 23 Å have been measured with nitroxide spin labels [221], PREs of up to 35 Å have been observed for chelated paramagnetic metals [222].

2.1.6.6 Measurement of diffusion by NMR spectroscopy

Diffusion NMR techniques provide a convenient way of differentiating between species in solution according to size and shape [223, 224]. The translational diffusion of a molecule is directly related to its hydrodynamic radius which in turn can be related (for folded proteins) to the rotational correlation time and molecular weight.

The diffusion coefficient can be experimentally measured using a pulsed gradient stimulated echo (PGSTE) NMR experiment. This type of NMR experiment relies on two gradients separated by time Δ to encode and decode the positions of the spins within the NMR sample (Fig2.9). The reversal of the encoding by the second gradient will only be perfect if spins have not diffused in the period between the two gradients. Diffusion leads to attenuation of the NMR signal and the dependence of this attenuation on the length of the period between the two gradients (Δ) and the strength (G) and length (δ) of the gradient field used in the encoding steps can be used to determine the diffusion coefficient according to the Stejskal-Tanner equation [225]:

$$I\left(G\right) = I_0 \exp\left[-G^2\delta^2 s^2\gamma^2\left(\Delta - \frac{\delta}{3}\right)D\right] \tag{2.1.16}$$

where $I(G)$ corresponds to the measured intensity at gradient strength G, I_0 is the initial intensity, D is the diffusion coefficient and γ the gyromagnetic ratio of the encoded/decoded nucleus. The hydrodynamic radius r_{h}, can readily be estimated from the experimentally derived diffusion coefficient D, by means of the Stokes-Einstein equation:

$$r_{\mathrm{h}} = \frac{k_{\mathrm{B}}T}{6\pi D\eta} \tag{2.1.17}$$

where k_{B} is the Boltzmann constant and η the solvent viscosity of water.

Two types of diffusion experiments are used throughout this study. The XSTE (heteronuclear stimulated echo) experiment permits the selection of specific chemical groups such that only protons attached to ^{15}N or ^{13}C probes are detected [226]. It relies on the

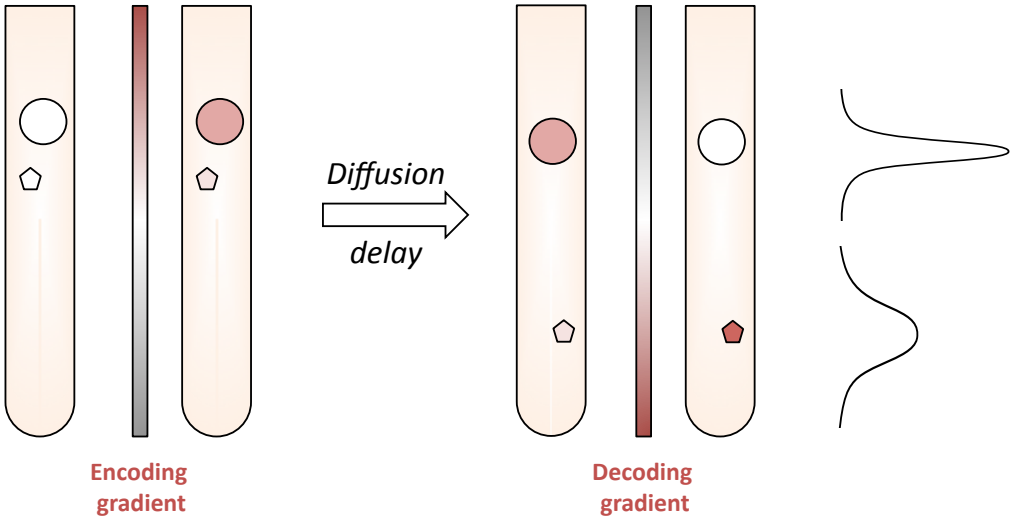

Figure 2.9: X-STE diffusion NMR. Diffusion NMR experiments rely on a set of gradient pulses to encode and decode the physical position of a particle in the NMR tube. The second, reverse gradient is applied after a diffusion delay. Large particles, which will have defused little during that period will have their encoding fully reversed and will give full intensity NMR signals. The signals of smaller particles will be reduced as the diffusion of the particle results in incomplete reversal of the encoding.

encoding of highly sensitive ^{1}H nuclei while heteronuclear editing reduces the complexity of the observed spectrum and permits the exploitation of favourable long relaxation rates of heteronuclei such as ^{15} and ^{13}C. A more recently developed sequence, the SORDID (signal optimisation with recovery in diffusion delays) experiment, optimises the diffusion and recycling delays (during which magnetisation returns to equilibrium) to rapidly repeat the diffusion experiment thus increasing the signal-to-noise acquired per unit time [227] by $\sqrt{2}$.

2.2 Protein expression and purification

2.2.1 Reagents

Standard lab chemicals, isotopes and D_2O were purchased from Sigma-Aldrich and Fisher Scientific. Metabolic precursors, keto-butyrate and isovalerate, were acquired from Isotec Isotopes.

2.2.2 Plasmids and Bacterial strains

Expression vectors carrying wild-type TF and cysteine mutants (positions 14, 150, 326 and 376) were kindly gifted by Prof Ulrich Hartl (Max Planck Institute of Biochemistry, Martinsried). TF cysteine mutant DNA sequences were preceded by a hexa-histadine tag and TEV-cleavage site while WT TF was expressed with a non-cleavable H_6-tag. A TEV protease cleavable WT TF clone (TF C14R) was prepared by site-directed mutagenesis (Section 2.2.3.1) from TF R14C template DNA.

Isolated α-synuclein was expressed from a pT7-7 vector, gifted by Prof Peter Lansbury (Harvard university). The α-synuclein sequence was sub-cloned by Dr Lisa Cabrita into the pLDC17 vector containing an N-terminal and C-terminal SecM stalling sequence using the *NheI* and *KpnI* restriction sites to generate the α-synuclein RNC construct. A stop codon was inserted before the SecM stalling sequence to generate the H_6-αsynuclein expression plasmid. A TEV protease cleavage site was introduce between the N-terminal H_6-tag and the α-synuclein DNA sequence by Annika Weise to generate the final α-synuclein RNC clone used in this study. Luciferase substitution mutant α-synuclein(87-100) constructs were prepared by site-directed mutagenesis by Anaïs Caissaignau.

ddFLN5 truncation protein clones (ddFLN5Δ16, ddFLN5Δ12, ddFLN5Δ6) were previously generated by Dr Marilia Karyadi. Full-length ddFLN5 and ddFLN5-6 were previously prepared by Dr Lisa Cabrita. Single tryptophan substitution clones were produced by site-directed mutagenesis from ddFLN5Δ16 template DNA by Millie Pang. The ddFLN5+21RNC constructs consists of an N-terminal his-tag and the ddFLN5 sequence directly followed by the SecM stalling sequence. This clone was previously generated by Dr Lisa Cabrita. All ddFLN5 constructs were carried on the pLDC vector and contained N-terminal his-tags for affinity chromatography.

All clones were under control of the isopropyl-β-D-1-thiogalactopyranoside (IPTG)-inducible T7 promoter and the plasmids carried genes for ampicillin resistance. All cloning was performed using *E. coli* DH5α cells. Expression of proteins was carried out in *E. coli* BL21*(DE3) Gold and BL21(DE3)Δtig cells to reduce background TF expression. All DNA and amino acid sequences can be found in Appendix A.1.

2.2.3 Growth media composition

The tables 2.1-2.3 summarise the components of typical growth media used in the preparation of 70S ribosomes, RNC and proteins. Variations on these recipes are required to accommodate ^{15}N ($^{15}NH_4CL$), ^{13}C (^{13}C glucose) isotope enrichment, deuteration (D7-glucose) or ILV labelling (precursors) and are described where relevant.

LB (Luria Bertani) medium (Table 2.1) is a rich complex media that was used for routine bacterial growths including overnight cultures, starter cultures, transformations, agar plates, unlabelled protein, ribosome and RNC preparations.

Enhanced M9 minimal medium (Table 2.2) was used for the expression of isotopically labelled and deuterated proteins and RNCs. *E. coli* cells for preparations of 70S ribosomes and RNCs were grown to high cell densities in MDG minimal media (Table 2.3).

Compound	**Concentration**
Tryptone	10 g/L
NaCl	10 g/L
Yeast extract	5 g/L

Table 2.1: LB media composition

Compound	**Stock**	**Concentration**
*EM*9 *salts* pH 8.0	10X	1x
Na_2HPO_4	71 g/L	25 mM
KH_2PO_4	34 g/L	25 mM
NaCl	58.4 g/L	50 mM
NH_4Cl	10 g/L	2 g/L
$MgSO_4$	1 M	2.5 mM
$CaCl_2$	1 M	100 μM
Glucose	20 g/L	2 g/L
BME vitamins (Sigma)	100 %	0.25 %
Trace metals	100 %	0.0125 %

Table 2.2: EM9 media composition.

Compound	Stock	Concentration
MDGSalts	50x	1x
NH_4Cl	2.5 M	50 mM
Na_2HPO_4	1.25 M	25 mM
KH_2PO_4	1.25 M	25 mM
Na_2SO_3	250 mM	5 mM
Glucose	20 g/L	4 g/L
L-aspartic acid pH 7.0	50 g/L	2 g/L
$MgSO_4$	1 M	2 mM
Trace metals [1]	100 %	0.2 %

[1] 50mM $FeCl_2$ (dissolved in 0.1M HCl), 20mM $CaCl_2$, 1mM $MnCl_2$, 1mM $ZnSO_4$, 2mM $CoCl_2$, 2mM $CuCl_2$, 2mM $NiCl_2$, 2mM Na_2MoO_4, 2mM Na_2SeO_3, 2mM H_3BO_3

Table 2.3: MDG media composition.

2.2.4 Molecular Biology

2.2.4.1 Site-directed mutagenesis

The QuickChange site-directed mutagenesis kit was used to create single amino acid substitution clones of TF (C14R), ddFNL5Δ16 (Y655W, F665W, A668W, F672W and E724W) and the 14 amino acid substitution of α-synuclein with luciferase residues 87-100.

Oligonucleotides : Oligonucleotide sequences used in the PCR reaction below, can be found in Appendix A. All oligonucleotides were purchased from Eurofins.

Polymerase chain reaction : DNA was amplified by polymerase chain reaction using a Corbin Thermal Cycler using KOD polymerase (Novagen) and the oligonucleotides above. The reaction mixture and temperature cycles used are detailed in Table 2.4 and 2.5 respectively.

PCR products were verified in first instance using 0.8 % (w/v) agarose gel electrophoresis (Section 2.2.4.5)

Reagent	Stock	Concentration	Volume
DNA template	100 ng/μL	10-500 ng	0.5 μL
5' primer	10 μM	150 nM	0.4 μL
3' primer	10 μM	150 nM	0.4 μL
dNTPs	2 mM	200 μM	2.5 μL
$MgSO_4$	1 mM	75 μM	1.5 μL
KOD Polymerase	3 u/μL	1.5 u	0.5 μL
Polymerase buffer	10x	1x	2.5 μL
ddH_2O	-	-	16.7 μL

Table 2.4: Typical polymerase chain reaction (PCR) composition

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	2 min	1
Denaturation	98 °C	10 sec	
Annealing	50-65 °C	30 sec	25
Extension	68 °C	2 min	
Final extension	72 °C	10 min	1

Table 2.5: Typical PCR temperature cycles employed in this study

2.2.4.2 DpnI digest

Remaining template DNA after PCR amplification was digested with *DpnI* by mixing 20 μL of PCR product, 2.5 μL and 1 μL of enzyme and 1.5 μL ddH_2O for a final reaction volume of 25 μL. The reaction was incubated at 37°C for 2 hours, then the enzyme was heat-inactivated at 65°C for 20 min.

2.2.4.3 Plasmid transformations

1 μL of plasmid was added to 25 μL of competent *E. coli* cells and incubated on ice for 30 min. The cell mixture was then heat-shocked at 42°C for 45 sec. The cells were then incubated on ice for 2 min before 800 μL of LB medium was added and the cultures were

incubating for 1 hour at 37°C shaking at 200 rpm for expression of ampicillin resistance genes. Cells were plated on LB-Amp plates and incubated at 37°C overnight.

2.2.4.4 DNA purification

DNA was purified from 5 ml cultures of a single colony of transformed DH5α cells grown overnight at 37°C using the QiaPrep Spin miniprep kit (Qiagen) following the manufacturer's instructions. New clones were then verified by DNA sequencing (Source Biosciences).

2.2.4.5 Agarose gel electrophoresis

PCR products were verified by agarose gel electrophoresis. 0.8% (w/v) agarose was dissolved by heating in 1X TAE buffer (50X stock: 242 g Tris, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA, pH 8.0). The solution was cooled to 45°C and 1 μg/ml ethidium bromide was added (from a 10 mg/ml stock) before the gel was cast. DNA samples were mixed 5:1 with loading dye (Promega) and loaded into the gel wells. Electrophoresis was performed at 100 V for ∼25 min. Gels were imaged under ultra-violet light.

2.2.5 Protein expression and purification

2.2.5.1 Protein over-expression and isotopic labelling

Isolated proteins were over-expressed in BL21(DE3)Δtig (TF) and BL21*(DE3) gold (α-synuclein, ddFLN5 and TEV protease) *E. coli* strains. Cells were transformed with over-expression plasmids as described in Section 2.2.3.3 and plated on LB-Amp agar plates, then grown overnight at 37°C. A single colony was used to inoculate a 5 ml pre-culture of LB medium with 0.1 mg/mL Ampicillin (LB-Amp). The pre-culture was incubated for 8 hours at 37°C shaking at 200 rpm. The pre-culture was then scaled up to 50 ml of LB -Amp which was incubated overnight at 37°C shaking at 200 rpm and cells grown to saturation.

Unlabelled proteins : This pre-culture was used to inoculate 500 ml to 1 L of LB-Amp medium to an optical density at 600 nm (OD_{600}) of 0.1. Cells were grown for 2-3 hours at 37°C with 200 rpm shaking. At an OD_{600} of 0.6, expression was induced with 1 mM of IPTG. Expression was allowed to proceed at 37°C for 4 hours (or 16-18 hours at 21 °C).

Uniform ^{15}N and ^{13}C isotopic labelling : The pre-culture was pelleted by centrifugation (3,000 rpm, Thermo Scientific 75003181 rotor, 20 min) and resuspended in 500 ml to 1 L of EM9 media (see Section 2.2.3) prepared with either $^{15}NH_4Cl$ (1 g/L) and/or ^{13}C glucose (2 g/L) for uniform ^{15}N and/or ^{13}C-labelling respectively. Cultures were typically

grown for 5-6 hours at 37°C with 200 rpm shaking until an OD_{600} value of 0.6 was reached, at which point protein expression was induced with 1 mM IPTG. The cultures were then incubated for another 4 hours at 37°C while shaking (or 16-18 hours at 21°C).

Perdeuteration and selective protonation of ile and I,L,V side-chain methyl groups : 5 ml of EM9 medium was inoculated with 1 ml of the pre-culture and grown overnight at 37°C while shaking at 200 rpm. This culture was used to inoculate 5 ml of EM9 medium (protonated components, 80% D_2O) and grown for 10 hours at 37°C. Cells were pelleted (3,000 rpm, Thermo Scientific 75003181 rotor, 20 min) and resuspended in 50 ml of EM9 medium (100% D_2O, protonated glucose) then grown overnight at 37°C. Finally, cells were pelleted and resuspended in fully deuterated EM9 media (100% D_2O, D-7 glucose). The culture was grown at 37° for 10-12 hours to an OD_{600} value of 0.5.The respective metabolic precursors with selectively protonated methyl groups for isoleucine (2-Ketobutyric acid-4-^{13}C,3-d_2, a.k.a keto-butyrate, Sigma-Aldrich) and leucine and valine (2-Keto-3(methyl-d_3)-butyric acid-4-^{13}C, 3-d_1, a.k.a isovalerate, Sigma-Aldrich) were added to 80 mg/L and the cultures grown for a further hour at 37°C. Over-expression was induced by 1 mM IPTG for 32 hours at 18°C.

All cultures were harvested by centrifugation at 4,000 rpm (Beckmann-Coulter JLA 8.1000 rotor) for 20 min. Cell pellets were transferred to 50 ml falcon tubes and stored at -20°C until purification.

2.2.5.2 Purification of trigger factor

Buffer composition

- Resuspension buffer : 25 mM Na_2HPO_4, 2 M NaCl, 2% Triton X-100, pH 8.0
- Ni-NTA wash buffer : 25 mM Na_2HPO_4, 500 mM NaCl, 25 mM Imidazole, pH 7.8
- Ni-NTA elution buffer : 25 mM Na_2HPO_4, 500 mM NaCl, 250 mM Imidazole, pH 7.8
- Size-exclusion buffer : 25 mM Na_2HPO_4, 100 mM NaCl, 2 mM BME, pH 7.8
- Tico buffer : 10 mM Hepes, 30 mM NH_4Cl, 12 mM $MgCl_2$, 5 mM EDTA, 2 mM BME, 0.1 % SigmaFast protease inhibitor cocktail tablet, pH 7.5

Purification procedure : Cell pellets were resuspended in 50 ml resuspension buffer per litre of culture, supplemented with lysozyme (100 mg/ 50 ml) and a SigmaFast protease inhibitor cocktail tablet. Cells were lysed by sonication for 10-15 cycles (30 s on and 30 s off) and the cellular debris was pelleted by centrifugation (18,000 rpm, 45 min, Sorvall SS-34 rotor). The lysate supernatant was then added to 10 ml of Ni-NTA resin

pre-equilibrated in resuspension buffer for His-tag affinity purification and incubated for 2 hours at 4°C. The flow-through was collected and the Ni-NTA resin washed to baseline with resuspension buffer as monitored via Bradford assay (Pierce, 160 μL reagent + 40 μL sample). A subsequent wash step to baseline in Ni-NTA wash buffer prevented non-specific interaction of proteins with the resin. TF was eluted in 50-100 ml Ni-NTA elution buffer and dialysed overnight into size-exclusion buffer. The dialysed fractions were then concentrated to 5 ml through a 10 kDa molecular weight cut-off (MWCO) concentrator (Millipore) for further purification by size-exclusion chromatography (SEC). The concentrated protein was applied to a superdex 200 column (S200) pre-equilibrated in size-exclusion buffer and eluted at a rate of 1 mL/min. SEC absorbance profiles at 280 nm and SDS-PAGE results were compared in order to asses sample purity. Highly pure TF fractions were concentrated and buffer-exchanged into H_2O or Tico buffer for further experiments.

Quantification and storage : TF concentrations were routinely determined by absorbance measurements at 280 nm using an extinction coefficient of 17,442 $M^{-1}cm^{-1}$ and cross-validated using the bicinchonic acid assay (BCA, Pierce) according to manufacture's protocol. Aliquots were flash-frozen in liquid nitrogen and stored at -80°C.

2.2.5.3 Additional purification steps

TEV-cleavage : The hexa-histadine (H_6) tag of TF (wild-type and cysteine mutants) was removed by cleavage (where mentioned) with TEV protease. This was done after the Ni-NTA step and involved dialysis overnight into size-exclusion buffer after addition of an optimised concentration of 1:10 (OD_{280} ratio) of TEV protease. TEV protease and uncleaved protein were then removed by a second Ni-NTA step. The TEV protease and TF mixture was applied to 10 ml pre-equilibrate Ni-NTA resin and incubated for 2 hours at 4°C. Cleaved TF was found in the flow-through, concentrated and purified further by size-exclusion chromatography as described Section1.2.4.2.

MTSL labelling : Electron paramagnetic resonance (EPR) and paramagnetic relaxation enhancement (PRE) experiments required the introduction of an 3-(2-Iodoacetamido)-PROXYL (proxyl) spin label at selected cysteine positions (14, 150, 326 and 376). After the Ni-NTA column, samples were incubated with 10 mM DTT for 4 hours at room-temperature at dilute protein concentrations to fully reduce accessible cysteines. The reducing agent was removed via size exclusion chromatography (in size-exclusion buffer but in the absence of BME). Then, the pooled fractions were diluted 2-3 fold to 20-50 μM protein concentration to avoid cross-linking of TF molcules, proxyl dissolved in DMSO

was added to 30 fold molar excess and incubated overnight at 4°C. Excess spin label was removed during the final SEC step of the purification as described above.

Ellman's assay : The proxyl labelling reaction was monitored by Ellman's assay. 50 μL sample was reacted with 10 μL Ellman's Reagent (Pierce) solution (4 mg/mL) in 500 ml reaction buffer (0.1 M $NaPO_4$, 1 mM EDTA, pH8.0) and incubated at room temperature for 15 min. The absorbance was measured at 412 nm and free cysteine concentrations calculated from the extinction coefficient (14,150 $M^{-1}cm^{-1}$). Typical spin-labelled TF samples were 90-100% labelled after overnight incubation with limited reactivity of reduced cysteines detected.

2.2.5.4 Purification of isolated α-synuclein and mutants

Buffer composition

- Lysis buffer: 100 mM Tris-HCl, 10 mM EDTA, 2 mM BME, Roche protease inhibitor tablet, pH 8.0
- Low salt buffer: 25 mM Tris-HCl, 2 mM BME, pH 7.8
- High salt buffer: 25 mM Tris-HCl, 1 M NaCl, 2 mM BME, pH 7.8
- Size-exclusion buffer: 25 mM Tris-HCl, 150 mM NaCl, 2 mM BME, pH 7.8

Purification procedure : Harvested cells were thawed and resuspended in 50 ml lysis buffer, then lysed by sonication for 10-15 cycles (30 s on and 30 s off). The cellular debris were pelleted by centrifugation (18,000 rpm, Sorvall SS34 rotor for 30 min). The supernatant was brought to the boil and left to simmer for 20 min, then, aggregates were pelleted by centrifugation (13,500 rpm, Sorvall SS34 rotor, 20 min). 10 mg/ml of streptomycin sulfate was then added in order to precipitate the remaining DNA/RNA and incubated for 30 min at 4°C, then, the sample was once again centrifuged (13,500 rpm, Sorvall SS34 rotor, 20 min). The remaining protein in the supernatant was precipitated by adding 400 mg/ml ammonium sulfate and incubating at 4°C for 30 min and pelleted by centrifugation (13,500 rpm, Sorvall SS34 rotor, 20 min). The pellet was resuspended in low salt buffer and dialysed O/N against H_2O. The resuspended protein was further purified by anion exchange. Proteins were eluted by a 60 ml 0-100 % high salt buffer gradient at 2.5 mL/min. α-synuclein containing fractions were pooled based on chromatography profile and SDS-PAGE results and concentrated to 5 ml using 3 kDa MWCO concentrators and applied to a S75 column for size-exclusion chromatography by isocratic elution (1 mL/min). Highly pure α-synuclein fractions were identified by SDS-PAGE, pooled, concentrated and buffer exchanged into H_2O or Tico buffer and stored at -80°C.

Quantification : α-synuclein protein concentrations were determined by absorbance measurements at 280 nm and calculated using an extinction coefficient of 5960 $M^{-1}cm^{-1}$. Calculated concentrations where cross-validated against BCA assay results as described Section 1.2.5.4.

2.2.5.5 Purification of isolated ddFLN5 and mutants

Buffer composition

- Resuspension buffer : 25 mM Na_2HPO_4, 2 M NaCl, 2 %(w/v) Triton X-100, pH 8.0
- Ni-NTA wash buffer : 25 mM Na_2HPO_4, 500 mM NaCl, 25 mM Imidazole, pH 7.8
- Ni-NTA elution buffer : 25 mM Na_2HPO_4, 500 mM NaCl, 250 mM Imidazole, pH 7.8
- Size-exclusion buffer : 25 mM Na_2HPO_4, 150 mM NaCl, 2 mM BME, pH 7.8
- Tico buffer : 10 mM Hepes, 30 mM NH_4Cl, 12 mM $MgCl_2$, 5 mM EDTA, 2 mM BME, 0.1% SigmaFast protease inhibitor cocktail tablet, pH 7.5

Purification procedure : Cell pellets were resuspended in 50 ml resuspension buffer supplemented with trace DNaseI, (100 mg/50 ml) lysozyme and a SigmaFast protease inhibitor cocktail tablet and lysed by sonication for 10-15 cycles (30 s on and 30 s off). The cellular debris was pelleted by centrifugation (18,000 rpm, Sorvall SS34 rotor 45 min) and the supernatant filtered (0.45 μm) before loading on a Ni-NTA column pre-equlibrated in resuspension buffer. The Ni-NTA column was washed to baseline with resuspension buffer as monitored via Bradford assay (described in Section 1.2.5.2) and then washed with Ni-NTA wash buffer, again to baseline determined by Bradford assay. ddFLN5 protein was eluted using Ni-NTA elution buffer and dialysed over-night in size-exclusion buffer. The protein was concentrated down to a volume of 5 ml using a 5 kDa cut-off concentrator and loaded on a pre-equilibrated Superdex S75 size-exclusion column. The protein was eluted by isocratic elution in size exclusion buffer. Fractions were collected based on the absorbance profile at 280 nm and SDS-PAGE results.

Quantification and storage : Purified ddFLN5 and mutants were buffer-exchanged into Tico buffer and concentrated to 500-1000 μM using a 5 kDa MWCO concentrator, aliquoted and flash-frozen in liquid nitrogen. Samples were stored at -80°C. Final ddFLN5 concentrations were determined from the absorbance at 280nm, using the coefficient value of 5960 $M^{-1}cm^{-1}$ as estimated from the amino acid sequence using the EXPASY prot-param tool [228].

2.2.5.6 Expression and purification of Tabacco Etch Virus (TEV) protease

Expression TEV-MBP fusion protein was expressed in *E.coli* strain BL21*(DE3) Gold from a single colony in auto-induction media, terrific broth (TB) (60 mg/L TB granules, 10 ml/L Glycerol) and grown overnight at 30°C until saturation (∼20 hours). Cultures were harvested by centrifugation (4,000 rpm, Beckmann-Coulter JLA 8.1000 rotor, 20 min) and stored at -20°C.

Buffer composition

- Lysis buffer : 25 mM Na_2HPO_4, 500 mM NaCl, 20%(v/v) Glycerol, 20 mM Imidazole, lysozyme (100 mg / 50 ml), a Roche protease inhibitor tablet, pH 7.5
- Ni-NTA wash buffer : 25 mM Na_2HPO_4, 500 mM NaCl, 20% Glycerol, 20 mM Imidazole, pH 7.8
- Ni-NTA elution buffer : 25 mM Na_2HPO_4, 500 mM NaCl, 20% Glycerol, 250 mM Imidazole, pH 7.8
- Dialysis buffer : 25 mM Na_2HPO_4, 200 mM NaCl, 20% Glycerol, 5 mM BME, pH 7.8
- Size-exclusion buffer : 25 mM Na_2HPO_4, 200 mM NaCl, 10% Glycerol, 5 mM BME, pH 7.8

Purification procedure : Cell pellets were resuspended in 50 ml lysis buffer, incubated at 4°C for 20 min and lysed by sonication for 10-15 cycles (30 s on and 30 s off). The cellular debris were pelleted by centrifugation (18,000 rpm, Sorvall SS34 rotor, 45 min) and the lysate was filtered (0.45 μm) before it was loaded onto a pre-equilibrated HisTrap column in lysis buffer. The column was washed with Ni-NTA wash buffer and the protein eluted in 50ml Ni-NTA elution buffer. The TEV protease was then dialysed overnight into dialysis buffer in two steps, then, concentrated using a 10 kDa MWCO concentrator to 5 ml or when aggregation is observed. Finally, the TEV protease was further purified by size-exclusion chromatography using a Sephadex S75 column pre-equilibrated in SEC buffer. TEV was eluted by isocratic elution at 1 ml/min in size exclusion buffer. Purified fractions were pooled based on OD_{280} reading and SDS-PAGE, concentrated to ∼5 mg/ml, flash-frozen and stored in aliquots at -80°C.

2.2.6 Growth and purification of 70S ribosomes

2.2.6.1 Growth

70S ribosomes were purified from *E. coli* BL21*(DE3) Δtig cells. A saturated pre-culture of 5 ml LB was used to inoculate a larger 50 ml LB starter culture grown for 10 hours at

37°C, which in turn was used to inoculate a final volume of 250 ml of MDG media. At each transition into fresh medium, cells were diluted to an OD_{600} value of 0.05 by pelleting the appropriate amount of culture by centrifugation (4,000 rpm, Thermo Scientific 75003181 rotor, 15 min) before being resuspended into fresh media. The final culture was grown at 30°C to an OD_{600} value of ∼3-6, reached after approximately 16 hours growth, before harvesting by pelleting for 15 min at 4,000 rpm (Beckmann-Coulter JLA 8.1000 rotor).

2.2.6.2 Purification

Buffer composition

- Lysis buffer: 50 mM Hepes, 1 M KOAc, 12 mM MgOAc, 2 mM BME, 5 mM EDTA, 5 mM adenosine triphosphate (ATP), 2 % Sigma protease inhibitor cocktail, 100 mg lysozyme, trace DNase, pH 7.5
- Cushion buffer: 50 mM Hepes, 1 M KOAc, 12 mM MgOAc, 2 mM BME, 5 mM EDTA, 5 mM ATP, 0.1 % Sigma protease inhibitor cocktail, pH 7.5
- Gradient buffer: 50 mM Hepes, 1 M KOAc, 12 mM MgOAc, 2 mM BME, 5 mM EDTA, 0.1 % Sigma protease inhibitor cocktail, pH 7.5
- Tico buffer: 10 mM Hepes, 30 mM NH_4Cl, 12 mM $MgCl_2$, 5 mM EDTA, 2 mM BME, 0.1% SigmaFast protease inhibitor cocktail tablet, pH 7.5

Purification procedure : Cells were resuspended in lysis buffer and lysed by French press (4 passes at ∼1000 psi), the cellular debris pelleted by centrifugation (18,000 rpm, Sorvall SS34 rotor, for 45 min) and the supernatant loaded on 35% sucrose cushions in cushion buffer, which was spun for 4 hours at 42,000 rpm in a Beckman Coulter Type 45 Ti rotor. The ribosome pellet was resuspended in 15 ml of Gradient buffer. Up to 1500 pmol of the obtained sample is loaded onto each of 6 sucrose gradients (35%-10%) in Gradient buffer. The gradients were spun overnight (15-16 hours) at 22,000 rpm using a Beckman Coulter SW28 rotor before fractionating; 70S ribosomes will have separated from 30S and 50S ribosome subunits by virtue of their size. Based on SDS-PAGE observation and fractionation absorbance profiles at 280 nm, fractions containing 70S ribosomes are pooled and concentrated using 100 kDa MWCO concentrator to a typical 500 μl sample of approximately 10 μM from each set of six gradients.

Quantification : Total ribosome concentrations were determined from absorbance measurements at 260 nm using an extinction coefficient of 4.2×10^7 $M^{-1}cm^{-1}$. A ratio of 1.9-2 of OD_{260}:OD_{280} indicated a highly pure ribosome sample.

2.2.7 Expression and purification of ribosome-nascent chain complexes

2.2.7.1 Expression

The ddFLN5+21 RNC was expressed in *E. coli* BL21*(DE3) Gold cells. A pre-culture of 5 ml LB grown overnight was used to inoculate a larger 100 ml LB starter culture grown for 10 hours at 37°C, which in turn was used to inoculate a final volume of 1 L of MDG media. At each transition into fresh medium, cells were diluted to an OD_{600} of 0.05 by pelleting the appropriate amount of culture by centrifugation (4,000 rpm, Thermo Scientific 75003181 rotor, 15 min) before being resuspended in the new media. This final volume was grown overnight (∼16 hours) at 30°C to saturation. Cells were pelleted by centrifugation 3,000 rpm (Thermo Scientific 75003181 rotor) for 25 min and washed twice in 1X EM9 salts (see Section 1.2.3) by repeatedly pelleting and resuspending the cells. Then, the cell pellet was resuspended in 1 L EM9 media and $^{15}NH_4Cl$ added. Expression was induced with 1 mM IPTG and the culture incubated at 30°C. 10 min after induction, 150 mg/L Rifampicin was added in order to block the *E.coli* RNA polymerase and expression allowed to proceed for 35 min before cells were harvested by centrifugation (3,500 rpm, Beckmann-Coulter JLA 8.1000 rotor, for 15 min).

2.2.7.2 Purification

Buffer composition

- Lysis buffer: 50 mM Hepes, 1 M KOAc, 12 mM MgOAc, 2 mM BME, 5 mM EDTA, 5 mM ATP, 2 % Sigma protease inhibitor cocktail, (100 mg/ 50ml) lysozyme, trace DNase, pH 7.5
- Cushion buffer: 50 mM Hepes, 1 M KOAc, 12 mM MgOAc, 2 mM BME, 5 mM EDTA, 5 mM ATP, 0.1 % Sigma protease inhibitor cocktail, pH 7.5
- Ni-IDA ATP wash buffer: 50 mM Hepes, 500 mM KOAc, 6 mM MgOAc, 2 mM BME, 5 mM ATP, 0.1 % Sigma protease inhibitor cocktail, pH 7.5
- Ni-IDA wash buffer: 50 mM Hepes, 500 mM KOAc, 6 mM MgOAc, 2 mM BME, 0.1 % Sigma protease inhibitor cocktail, pH 7.5
- Ni-IDA elution buffer: 50 mM Hepes, 500 mM KOAc, 6 mM MgOAc, 2 mM BME, 0.1 % Sigma protease inhibitor cocktail, 150 mM Imidazole, pH 7.5
- Gradient buffer: 50 mM Hepes, 1 M KOAc, 12 mM MgOAc, 2 mM BME, 5 mM EDTA, 0.1 % Sigma protease inhibitor cocktail, pH 7.5
- Tico buffer: 10 mM Hepes, 30 mM NH_4Cl, 12 mM $MgCl_2$, 5 mM EDTA, 2 mM BME, 0.1% SigmaFast protease inhibitor cocktail tablet, pH 7.5

Purification procedure Cell were resuspended in lysis buffer and lysed by French press (4 passes at ∼1000 psi), the cellular debris pelleted by centrifugation (18,000 rpm, Sorvall SS34 rotor, for 45 min) and the supernatant loaded on 35% sucrose cushions in Cushion buffer, which was spun for 4 hours at 42,000 rpm in a Beckman Coulter Type 45 Ti rotor. The sucrose was discarded and the ribosome pellet resuspended in Ni-IDA ATP wash buffer. The resuspended ribosomes were mixed with Ni-IDA beads and incubated for 2 hours at 4°C. The Ni-IDA beads were first washed to baseline (as monitored by Bradford assay, see Section 2.2.5.2) in Ni-IDA ATP wash buffer, then Ni-IDA wash buffer. RNCs were eluted from the Ni-IDA beads with 50 ml of Ni-IDA elution buffer and rapidly diluted with 1:1 gradient buffer. The total ribosome concentration was measured and up to 1500 pmol was loaded onto each of 6 sucrose gradients (35%-10%) in gradient buffer. The gradients were spun overnight (15-16 hours) at 22,000 rpm (Beckmann Coulter SW28 rotor) before fractionating. Fractions were pooled based on SDS-Page results and fractionation absorbance profiles and concentrated to a 500 μl sample of approximately 10 μM from each set of six gradients.

Quantification Total RNC concentration were determined from absorbance measurements at 260nm using an extinction coefficient of 4.2x10^7 $M^{-1}cm^{-1}$. A ratio of 1.9-2 of OD_{260}:OD_{280} indicated sufficiently pure RNC samples. Nascent chain occupancy was determined by Western blot and densitometry analysis against a known protein standard (Section 2.2.8.2).

2.2.8 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and immunodetection.

2.2.8.1 SDS-PAGE

Buffer composition

- Running buffer : 1x MOPS-SDS running buffer (Alfa Aesar)
- Coomassie stain: 40% ethanol, 10% glacial acetic acid, 0.05% Brilliant blue R-250 (Fisher BioReagents)
- Destain solution: 40% ethanol, 10% glacial acetic acid

Experimental procedure : Sodium dodecyl sulfate (SDS) polyacrylamide gels were cast according to the recipe in Table N. Running gels were cast first and the topped by a 0.5-1 cm stacking gel.

Reagent	Stacking	12%	15%
1.25 M Bis-Tris	1 mL	2.86 mL	2.86 mL
30% Acrylamide	0.53 mL	4 mL	5 mL
25% APS	9 μL	100 μL	100 μL
TEMED	9 μL	4 μL	4 μL
H_2O	2.47 mL	3.14 mL	2.14 mL
Total volume	5 mL	10 mL	10 mL

Table 2.6: SDS-PAGE gel composition

Running gels of 12% were used to visualise ribosome, RNC and TF preparations while 15% gels were used in the preparations of isolated α-synuclein and ddFLN5 proteins.

SDS-PAGE gels were run at 200 V for 35 min and were stained with Coomassie stain for protein preparation and with Silver stain (SilverQuest, Invitrogen) for ribosome and RNC purifications following manufacturer's instructions.

2.2.8.2 Imunnodetection

Buffer composition

- Transfer buffer: 6 g/L Tris, 14.4 g/L glycine, 20% (v/v) methanol, 0.01% (w/v) SDS
- Tris-buffers Saline (TBS): 2.42 g/L Tris, 11.69 g/L NaCL, pH 7.4
- TBS-Tween (TBST): 0.05%(v/v) Tween-20 in TBS
- Blocking buffer:

  Anti-H_6 blocking buffer : 1%(w/v) casein in TBS
  Anti-TF blocking buffer: 0.5%(w/v) milk powder in TBST

- Antibodies

  Anti-H_6 detection: Penta-His HRP conjugate (1:2500) (Qiagen) in TBST
  Anti-TF primary antibody: rabbit anti-TF polyclonal antibody (1:2500) (Genscript) in blocking buffer
  Anti-TF secondary antibody : anti-rabbit HRP-conjugate (1:1000)(New england Biolabs) in blocking buffer

- Chemiluminescence detection: SuperSignal West Pico (Femto where necessary) chemiluminescence substrate (Pierce)

Western blot procedure The SDS-PAGE gel was soaked in cooled transfer buffer and stacked into the transfer chamber with nitrocellulose membrane. Transfer was undertaken at 250 mA for 2 hours at room temperature. After transfer, the nitrocellulose membrane was soaked in blocking buffer while shaking. The gel was stained with Coomassie stain to verify efficient transfer has occurred. The membrane was then incubated with primary antibody for 2 hours (room temperature) or overnight at 4°C. The gel was then washed 4 times for 10 min in TBST. Where necessary, the membrane was then incubated with the secondary antibody for 1 hour at room temperature followed by further washing (4x 10 min in TBST). The membrane was incubated with chemiluminescence substrate and detected using a Fujifilm LAS-1000 scanner with exposures ranging from 30 s to 5 min depending on chemiluminescence intensities.

Densitometry analysis The intensities of the western blot signal were quantified by densitometry using ImageJ software using known protein standards.

2.3 NMR data acquisition and analysis

2.3.1 Experimental conditions

NMR experiments were recorded on a Bruker Advance III 700MHz spectrometer, Bruker Advance III HD 800MHz and Bruker Advance III HD 900MHz all equipped with cryo-probes and a Bruker Advance III 500MHz spectrometer. All experiments were recorded in Tico buffer (10 mM Hepes, 30 mM NH_4Cl, 12 mM $MgCl_2$, 5 mM EDTA, 2 mM BME, 0.1% SigmaFast protease inhibitor cocktail tablet, pH 7.5) supplemented with 10% D_2O as a lock solvent and 0.001% DSS, unless stated otherwise. NMR experiments of TF in the absence of substrates were acquired at 25°C in 99% D_2O while experiments involving α-synuclein and ddFLN5 were acquired at 4°C and 10°C respectively.

2.3.2 NMR data collection

2.3.2.1 1D proton spectra

^{1}H 1D experiments ^{1}H 1D spectra were acquired using excitation sculpting (*zgesgp*) for water-suppression with a spectral width of 20 ppm and 16384 time-points for an acquisition times of ca. 600ms. A recycle delay of 1 s was used between scans.

Heteronuclar-edited 1D experiments : ^{13}C- and ^{15}N-edited 1D experiments were recorded by acquiring the first increments of the 2D ^{13}C-heteronuclear mulitple-quantum coherence (HMQC), ^{15}N- heteronuclear single-quantum coherence (HSQC) and ^{15}N-band-selective optimised-flip-angle short-transient (SOFAST)-HMQC sequences (Sections 2.3.2.2, 2.3.2.3 and 2.3.2.4).

Isotopic labelling experiments The ^{15}N background labelling of ribosomes within RNC samples was determined using a pair of modified, single increment,^{15}N-SOFAST-HMQC experiments. A pre-saturation pulse was added to the 1D ^{15}N-edited experiment to improve water-suppression. The second ^{15}N-filtered experiment was run using the same sequence with the phase-cycle of the receiver inverted to reject ^{15}N-labelled magnetisation. The sum of the signal in both spectra corresponds to 100% of the ribosome signal, while the fraction of the ^{15}N-edited experiment signal over this summed total intensity corresponds to the fraction of ^{15}N labelled ribosomes in the sample.

2.3.2.2 ^{1}H-^{13}C HMQC correlation spectra

^{1}H-^{13}C-HMQC spectra (bruker library pulse sequence: *hmqcphpr*) of TF^{ile} and TF^{ILV} were typically recorded with 4096 points and a spectral width of 16 ppm in the direct dimension and 512 points and a spectral width of 22 ppm centred at 16 ppm in the

indirect dimension for acquisition times of 180 ms and 70 ms for ^{1}H and ^{13}C dimensions respectively. A recycle delay of 1s was used between scans.

2.3.2.3 ^{1}H-^{15}N HSQC

For α-synuclein and mutants, sensitivity enhanced ^{1}H-^{15}N HSQC correlation spectra (bruker library pulse sequence: *hsqcetfpf3gpsi2*) were recorded with 2048 points and a spectral width of 15 ppm in the direct dimension and 256 points and a spectral width of 26 ppm in the indirect dimension for acquisition times of ca 100 ms and 70 ms for ^{1}H and ^{15}N respectively. A recycle delay of 1 s was used between scans.

In the case of ddFLN5 and mutants, slightly lower resolution spectra were recorded with 128 point in the ^{15}N dimension of 26 ppm for a total acquisition time of 20ms.

2.3.2.4 ^{1}H-^{15}N SOFAST HMQC

^{1}H-^{15}N SOFAST HMQC spectra (*sfhmqcf3gpph*) [229] were recorded with 1024 points and a spectral width of 16 ppm in the direct dimension and 128 points and a spectral width of 32 ppm in the indirect dimension for acquisition times of ca 45 ms and 30 ms for ^{1}H and ^{15}N respectively. The recycle delay between scans was set to 50 ms.

2.3.2.5 Diffusion

STE diffusion experiments ^{1}H diffusion experiments of isolated protein species used stimulated gradient echos (STE, Section 2.1.6.6) where the diffusion delay (Δ = 100 ms) and gradient lengths (δ= 4 ms) were fixed and the gradient strengths varied up to 95% maximum gradient strength (G_{max} = 55.57 G/cm). 1D ^{1}H STE diffusion experiments were acquired with 2048 points over 20 ppm spectral width for an acquisition time of ca. 75 ms. 16 gradient strength points evenly distributed from 5% to 95% G_{max}. A recycle delay of 1s was used between scans.

SORDID diffusion experiments ^{15}N SORDID (Signal Optimisation with Recovery in Diffusion Delays) diffusion experiments [227] were acquired in order to monitor attachment of nascent chains to the ribosome. A diffusion delay of 190 ms and a gradient length of 2 ms were used. Two gradient strengths were used: 15% and 100% of 38.6 G/cm. Spectra were recorded with 1024 points over 15 ppm spectral width and a ^{1}H acquisition time of 50 ms.

2.3.2.6 Assignment via triple-resonance NMR experiments

Backbone NMR assignment data for α-synuclein(Luc 87-100) was acquired using the BEST (band-selective excitation short-transient) implementations of the standard HNCO, HN-

caCO, HNCACB and HNcoCACB triple-resonance assignment experiments [230]. The HNCO/HNcaCO experiments were acquired with 1024x84x96 points in the ^{1}H, ^{15}N and ^{13}C dimensions over spectral widths of 14 ppm, 32 ppm and 11 ppm for acquisition times of 50 ms, 25ms and 30ms. HNCACB/HNcoCACB experiments were acquired with 1024x84x192 points in the ^{1}H, ^{15}N and^{13}C dimensions over spectral widths of 14 ppm, 32 ppm and 64 ppm for acquisition times of 50 ms, 25 ms and 10 ms respectively. Recycle delays between scans were 300 ms long.

2.3.2.7 Cross-saturation experiments of TF

80 μM of TF^{ILV} was mixed with 320 μM non-isotopically labelled TF in 100 % deuterated Tico buffer (pH 7.5) with 0.001% DSS at 25°C. A standard HMQC pulse sequence was modified to include a 1.2 s delay period during which a selective continuous wave saturation pulse could be applied at 4.13 ppm. The experiment was recorded with and without the saturation pulse in an interleaved manner. Spectra were acquired with 4096 points over 16 ppmin the ^{1}H dimension and 256 points over 22 ppm centred at 16ppm in the ^{13}C dimension for total acquisition times of 180 ms and 35 ms respectively.

2.3.2.8 Relaxation dispersion of TF

CPMG relaxation experiments [182] were recorded on 200-800 μM TF^{ILV} (typical 600 μM) in 100% deuterated Tico (pH 7.5) at temperature of 15°C, 20°C and 25°C for magnetic field strengths of 700 and 900 MHz while data at 25°C only was acquired for 500 and 800 MHz. Multiple quantum relaxation dispersion was recorded using a TROSY optimised pulse sequence as described in [182]. CPMG frequencies used were 0, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000 Hz under all experimental conditions and the selective ^{1}H 180° inversion pulse calibrated and centred at 0.7 ppm. The total CPMG delay was 20 ms while the recycling delay between scans was set to 1.5 s. Spectra were acquired with 4096 points in the direct dimension with a spectral width of 16 ppm and 256-512 points over 22 ppm in the indirect ^{13}C dimension for acquisition times of 120 and 45 ms respectively.

2.3.2.9 Order parameters

^{1}H transverse relaxation of 600 μM TF^{ILV} at 25°C and 800MHz was measured according to [211]. Double-quantum coherence build-up was recorded using the same sequence with an additional 90° pulse after initial preparation of transverse magnetisation. Relaxation delays were 0.7, 1, 1.5, 2, 3, 4, 5, 7, 9, 12, 15, 20 ms. Spectra were acquired with 4096 points in the direct dimension with a spectral width of 16 ppm and 512 points over 22

ppm in the indirect ^{13}C dimension for acquisition times of 180 and 70 ms respectively. The recycle delay between scans was 1s.

2.3.3 NMR processing and analysis

DSS referencing In all spectra, ^{1}H dimensions were directly referenced to the internal reference DSS [190]. ^{15}N and ^{13}C chemical shifts were then referenced indirectly using their gyromagnetic ratios γ_N/γ_H=0.10132912 and γ_C/γ_H=0.25144953.

FID processing All NMR spectra were processed in nmrPipe [231]. Zero-filling, typically to two fold was applied to the FID. For NMR spectra of the RNC, an exponential window function was applied with line-broadening of 12 Hz in the ^{1}H dimension. A Lorentz-to-Gauss window function was applied to relaxation data of TF^{ILV} with an inverse exponential width of 8 Hz and 24 Hz gaussian line-broadening. A sine-bell window function with an offset of 0.5 and an exponent of 0.5 was applied to 1D and titration data. Data were converted to the time domain by Fourier transform and baseline correction applied where necessary (sample scripts can be found in Appendix A.2).

Titration data Processed 2D spectra were imported to Sparky [232] for comparison, transfer of assignments and peak height and chemical shift extraction. Heights and chemical shift changes were compared using the Excel (Microsoft) and Matlab (Mathworks) software packages. Titration data were fitted in Matlab by least-square minimisation of the numerical solution to the equilibrium as described by the equilibrium constants (see Sections 2.4 and 3.4).

NMR backbone assignment data Assignment of α-synuclein(Luc 87-100) was undertaken in CCPNMR Analysis [233].

NMR diffusion data NMR diffusion data was imported to Matlab, where resonance signals were integrated and fitted to the Stokes-Einstein equation (Section 2.1.6.6).

Relaxation data Relaxation data over the various relaxation delays and CPMG frequency were fitted by least-square minimisation in Fuda. Order parameters were determined by fitting the ratio of heights of the transverse relaxation and double-quantum coherence experiments in Matlab as described Section 2.1.6.3, equation 2.1.10. The intramethyl ^{1}H-^{1}H dipolar cross-correlated relaxation rates were then used to calculate the S^2 of each residue according to equation 2.1.9 . The fitted heights of the CPMG data were further imported into GUARDD [234] and fitted to equations 1.1.1-1.1.5 in order to determine kinetic parameters of observed exchange events.

2.4 Mathematical modelling of 5-state TF equilibrium

The complex TF equilibrium described in detail in Sections 1.2.2.5 and 3.4 was the subject of mathematical modelling in order to allow predictions of the behaviour of the TF populations under various experimental conditions. The various states are described by the equilibrium constants and numerically solved with the aid of the Mathematica 8 software package (Wolfram, version 8.0.4.0). Unless stated otherwise, predictions made about the TF population distribution under specific experimental conditions assumed values for the dissociation constants for dimerisation, ribosome binding and RNC binding of 1.35 μM, 1 μM and 100 nM respectively based on experimentally derived values in the literature more extensively discussed Section 3.4.

2.5 Electron Paramagnetic Resonance (EPR)

2.5.1 Theoretical basis of EPR

The fundamental principles of EPR spectroscopy are analogous to those of NMR, as outlined in Section 2.1. A primary difference is that the magnetisation of a lone electron spin is investigated. Electron magnetisation is significantly stronger than that of nuclei (660 fold) but also relaxes much faster. While continuous wave (cw) EPR experiments can be undertaken at room temperature, pulsed EPR experiments, where electron magnetisation is manipulated, are undertaken at 50 K to slow the rates of relaxation mechanisms.

Again, electron magnetisation returns to equilibrium according to similar relaxation mechanism as nuclear magnetisation (Section 2.1.2). Most contributions to transverse relaxation can be refocussed using a spin echo sequence but the coupling of two nearby electrons remains. The amplitude of the transverse magnetisation as a function of time is detected by DEER (double electron-electron resonance) experiments [235] thus informs on the proximity and orientation of nearby electrons. Since free electrons are typically not stable and don't naturally occur in biological macromolecules aside from paramagnetic metal ions, free electrons can be selectively introduced in a required position using spin-label or metal-binding tags in a controlled manner.

The observed oscillation in the transverse magnetisation decay corresponds to the dipolar coupling, ν_{dd}, which is described by :

$$\nu_{dd} = \frac{\mu_0 \hbar \gamma_A \gamma_B}{8\pi^2 r_{AB}^3} \left(1 - 3\cos^2\theta\right) \tag{2.5.1}$$

where, μ_0 is the vacuum permeability, $\hbar$ is the reduced Planck constant, γ_A and γ_B are the gyromagnetic ratios of the two electron spins, r_{AB} is the distance between them and θ is the angle between the distance vector r_{AB} and the external magnetic field B_0.

At least one oscillation is required to accurately determine ν_{dd} and thus the distance between the two electrons. DEER measurements are therefore limited to distances between 1.5 and 8 nm [235].

2.5.2 Experimental conditions

DEER experiments of TF were undertaken at 300 μM TF monomer concentration in deuterated Tico (10 mM Hepes, 30 mM NH_4Cl, 12 mM $MgCl_2$, 5 mM EDTA, 2 mM BME, 0.1% SigmaFast protease inhibitor cocktail tablet, pH 7.5) with 50 % deuterated glycerol. α-syn 87-100 up to 750 μM was added from a 2 mM stock to minimise dilution. Proxyl spin-label attachment to TF was verified by cw EPR at room-temperature before the sample was flash frozen. DEER experiments were performed at 50 K on a Bruker ELEXSYS E580 spectrometer operating at 9.6 GHz equipped with an ER 4118X-MS2 resonator, Oxford Instruments continuous flow cryostat (CF935) and ITC503 temperature controller. The four-pulse DEER sequence used was $\pi/2(\nu_{obs})$-τ_1-$\pi(\nu_{obs})$-t'-$\pi(\nu_{pump})$-(τ_1+τ_2-t')-$\pi(\nu_{obs})$-τ_2-echo, where the observer pulse length was 16 ns for $\pi/2$ and 32 ns for π pulses. The pump pulse length was 12 ns. The long inter-pulse delay (τ_2) was 3,500 ns. All other parameters were according to [236] with $\tau_{1,0}$=400 ns and $\Delta\tau_1$=56 ns. Data points were collected in 8 ns time steps. The spectra were analysed using the programme DEERAnalysis2013 [237]. The background was corrected by a homology three-dimensional fit and the distance distributions evaluated by Tikhonov regularisation.

Chapter 3

Investigation of the isolated trigger factor chaperone by NMR spectroscopy

3.1 Introduction

In isolation, trigger factor (TF) exists in a dynamic equilibrium between a monomer and dimer state which assembles with low μM affinity [148, 159, 160]. The characterisation of TF in the absence of protein substrates thus far has in large part focussed on the structure of the TF monomer [84, 126, 238], in order to provide a high resolution model for the interpretation of biochemical and biophysical observations in the presence of nascent chains. Where most studies of the TF interaction with RNCs are undertaken at low concentrations avoiding the dimeric state, the sensitivity limitations of NMR mean that such conditions are not accessible. Indeed, at concentrations suitable to NMR (>10 μM), the TF population is predicted to be largely dimeric. As NMR signals are the populated weighted average of the various species in solution, an improved understanding of the TF dimerisation reaction and its NMR properties is required for the interpretation of NMR data in the more complex nascent chain systems. Moreover, the structural data of the TF dimer and FRET restraint models [84, 126, 148, 238] disagree on fundamental properties such as the orientation of monomers and buried surfaces, highlighting a gap in our current understanding of this simple TF equilibrium.

In this Chapter, the possibility of investigating the TF chaperone by NMR spectroscopy is explored, using a selective protonation-based isotopic-labelling protocol in combination with methyl TROSY-optimised experiments. This methodology is then successfully applied to the determination of kinetic, thermodynamic and structural parameters of the TF dimerisation equilibrium. Mathematical models are developed for the interpretation

of these result in the context of the complex TF equilibrium.

3.2 Methods

Selective ^{1}H-^{13}C methyl Ile (TF^{ile}) and ILV (TF^{ILV}) labelled TF was prepared as described in Section 2.2.5.2 of the Materials and Methods. Mathematical modelling of the TF equilibria as described in Section 3.4 was undertaken using Mathematica 8 (Wolfram Research, 8.0.4.0).

3.3 Selective ^{1}H-^{13}C methyl labelling of Ile, Leu and Val residues in perdeuterated TF.

Established biomolecular NMR techniques rely for sensitivity on the enrichment of heterologously expressed proteins with magnetic isotopes such as the commonly used ^{13}C and ^{15}N nuclei. As described in Section 2.1.4, these isotopic labelling strategies have permitted protein studies by enhancing the NMR signal respective to a natural abundance sample and the simplification of spectra by multi-dimensional projections. However, the slow correlation times leading to rapid relaxation of the NMR signal and spectral crowding that characterise large macromolecules and complexes leave them inaccessible to uniform isotopic labelling strategies (Section 2.1.2).

Over the last decade, the development of methyl-TROSY experiments (described in Section 2.1.3) have permitted the exploitation of the inherent slow relaxation rates of methyl groups within amino acids to selectively increase the observed signals for these atoms on a perdeuterated background (reviewed in [239]). An *in vivo* expression protocol based on synthetic, labelled and deuterated, metabolic precursors for valine/leucine, isoleucine and more recently methionine and alanine residues has been developed for the selective incorporation of ^{1}H-^{13}C to the extreme (β, γ or δ) methyl groups [200]. Moreover, expression cultures are grown in highly deuterated growth media resulting in >99% perdeuteration of the over-expressed protein. Sample perdeuteration suppresses background signals from proton-coupled natural abundance carbon 13 signals elsewhere in the protein but more importantly dramatically decreases transverse relaxation rates by reducing the proton density in the protein environment. Additionally, methyl group containing amino acids have above average hydrophobicity localising them typically to folded cores of proteins making them excellent reporters of structural integrity and changes in proteins [203, 204]. In this study, the selective protonation of isoleucine and leucine / valine residues of TF is undertaken in order to improve the quality of the dimeric TF NMR spectra.

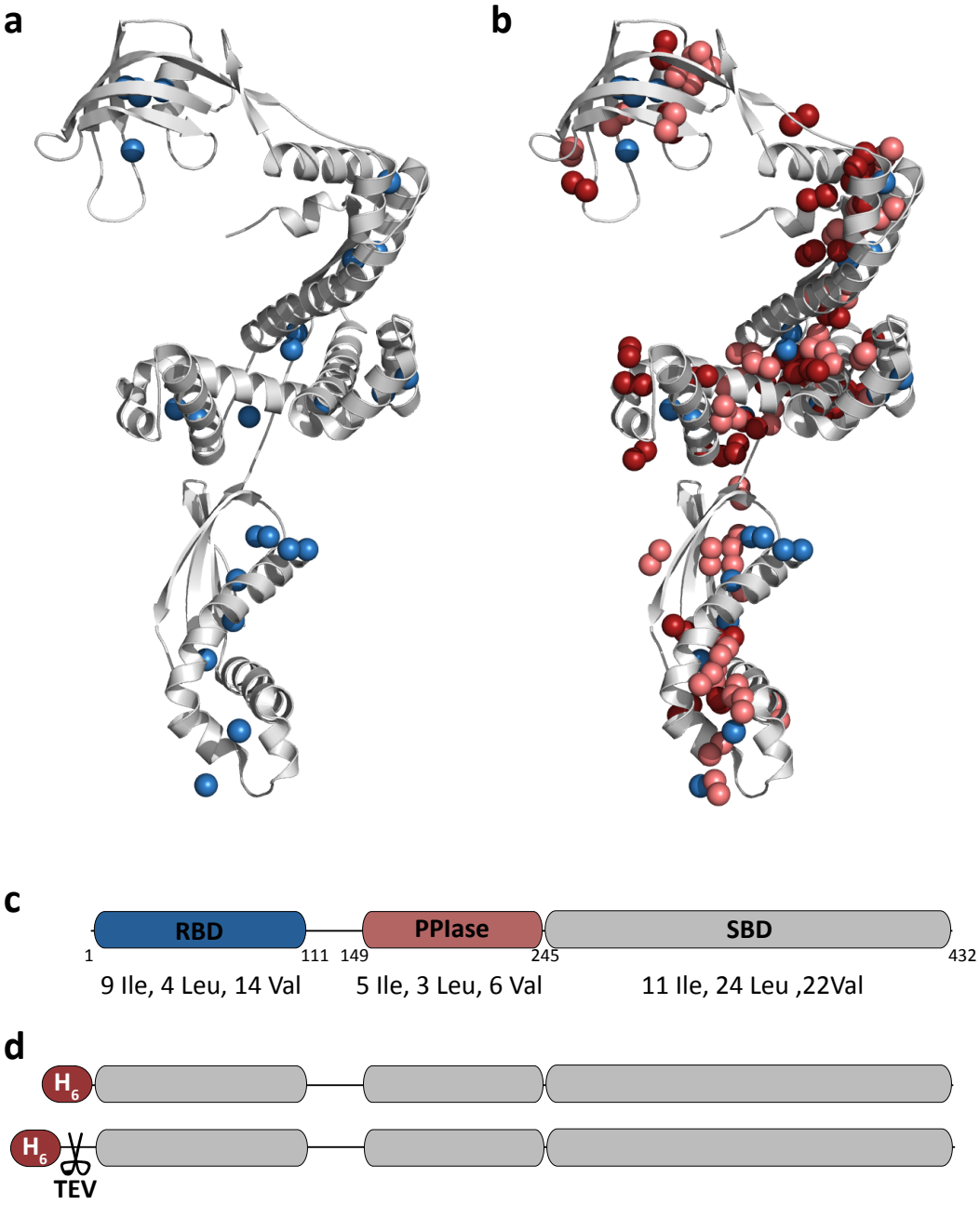

Figure 3.1: Trigger factor constructs and NMR probes. (**a**) Structural distribution of the 25 native TF isoleucine residues (side-chain methyl groups in blue). (**b**) Structural distribution of isoleucine (blue), leucine (red) and valine (pink) methyl side-chain groups. (**c**) Ile, Leu and Val residues are evenly distributed across the TF domains and represent almost 25% of the amino acid sequence. (**d**) H_6-tagged and TEV protease cleavable full-length TF constructs were used through-out this study.

3.3.1 Strategy for the selective ^{1}H-^{13}C methyl labelling of TF isoleucine residues.

The *E. coli* trigger factor contains 25 isoleucine residues (Fig 3.1, TF primary sequence in Appendix A.1.1) which were selectively labelled using the procedure outlined in the Materials and Methods (Section 2.2.5.1). In brief, *E. coli* strain BL21(DE3)Δtig cells are transformed with an expression vector containing the TF DNA sequence preceded by a N-terminal hexa-histadine (H_6-tag) purification tag and a TEV protease cleavage site and grown in deuterated M9 growth medium enriched with α-ketobutyrate, a metabolic precursor in isoleucine synthesis [200].

The growth procedure relies on the step-wise adaptation of cultured cells to growth in D_2O (Fig 3.2). Cells most able to adapt essential cellular functions to accommodate the heavy ^{2}H isotopes are evolutionary favoured by this approach without exerting the selective pressure as to cause expression plasmid loss. *E. coli* cells appear particularly sensitive to growth in D7-glucose, the deuterated carbon source. The effect of D7-glucose on cellular function is reflected in the results of the TF expression test in Figure 3.2. The western blot results clearly show delayed peaking of expression in the presence of D7-glucose. Expression in LB and deuterated M9 peaks at 16-20 hrs over-night induction at 18°C while expression in M9 with D_2O and D7-glucose is maximum at 24 hrs. Slower induction at reduced temperatures (18°C) appear to provide more favourable conditions to over-expression of TF in both LB and D_2O over more typical short (4hr) expression periods at 37°C (Fig 3.2, lanes 2-4).

3.3.2 Preparation and biochemical characterisation of TF^{Ile}

Preparation of TF : TF^{Ile} was prepared according to the purification protocol detailed in Section 2.2.5.2 and summarised in Fig 3.3.a. The initial Ni-NTA affinity chromatography step permits the enrichment of the lysate in H_6-TF. Where required, H_6-tags were cleaved with TEV protease. The TEV protease cleavage reaction was optimised for concentration and duration to 1:10 OD_{280} for 2 hours at room-temperature (Fig 3.3.d) or overnight at 4°C. A second Ni-NTA step then allowed the separation of cleaved TF from uncleaved H_6-TF, TEV and Ni-binding contaminants. While TF samples were highly pure at this stage (Fig 3.3.b lane 8), an additional SEC permitted the removal of remaining trace amount of TEV protease for fully purified TF samples (Fig 3.3.c). A typical deuterated preparation yielded 240 nmol of TF, or 300μL (one NMR sample) at 800μM. Deuterated NMR buffers were supplemented by protease inhibitors (0.1% w/v) for additional stability during experiment acquisition.

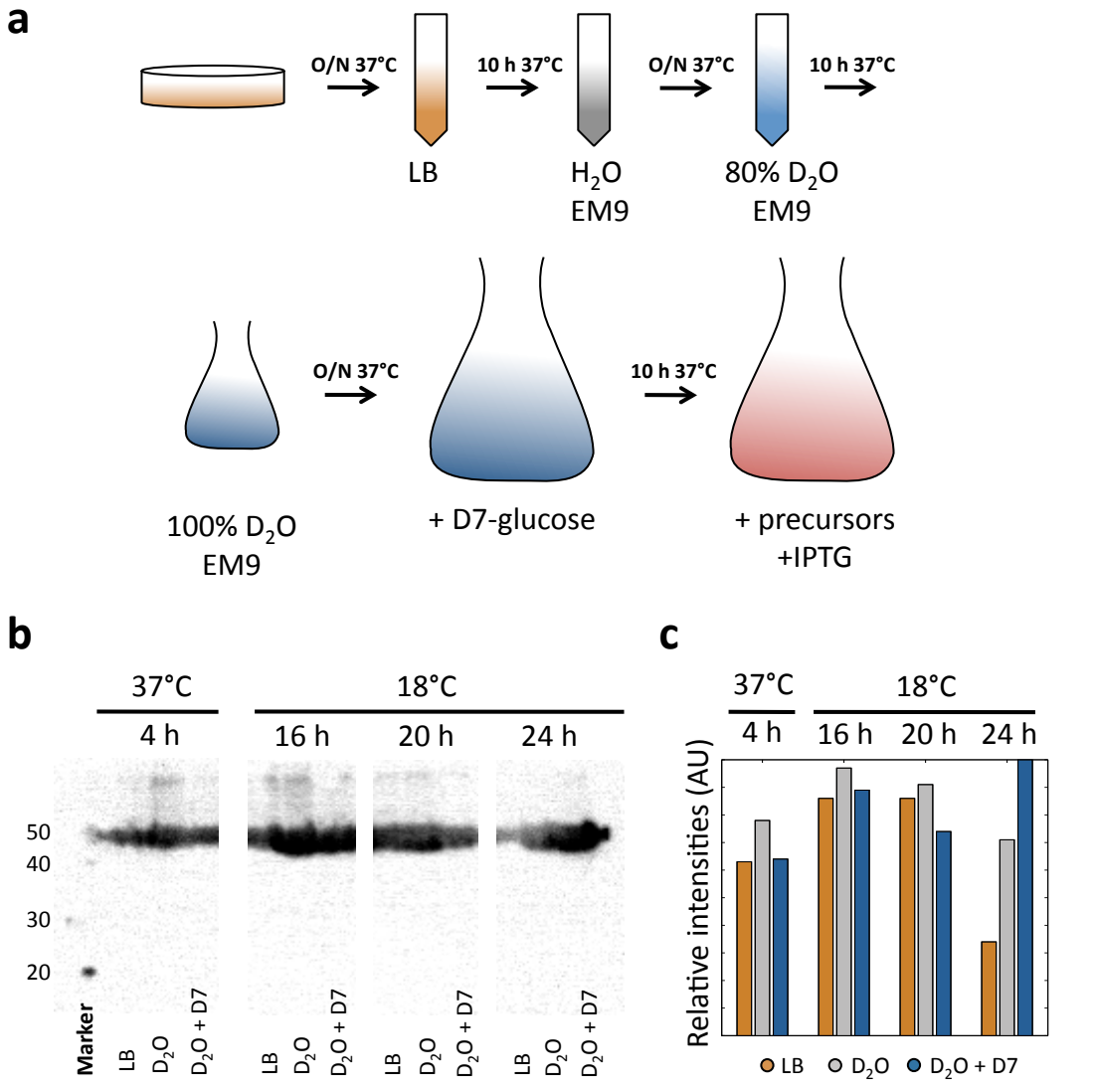

Figure 3.2: Trigger factor ILV labelling procedure. (**a**) Growth procedure for perdeuteration and selective ^{1}H-^{13}C labelling of side-chain methyl groups of over-expressed TF. Refer to text (Section 3.3.1 and 2.2.5.1) for details. (**b**) Expression test of TF in various expression media (5 ml cultures) and expression conditions. (**c**) Quantification of expression test results by densitometry.

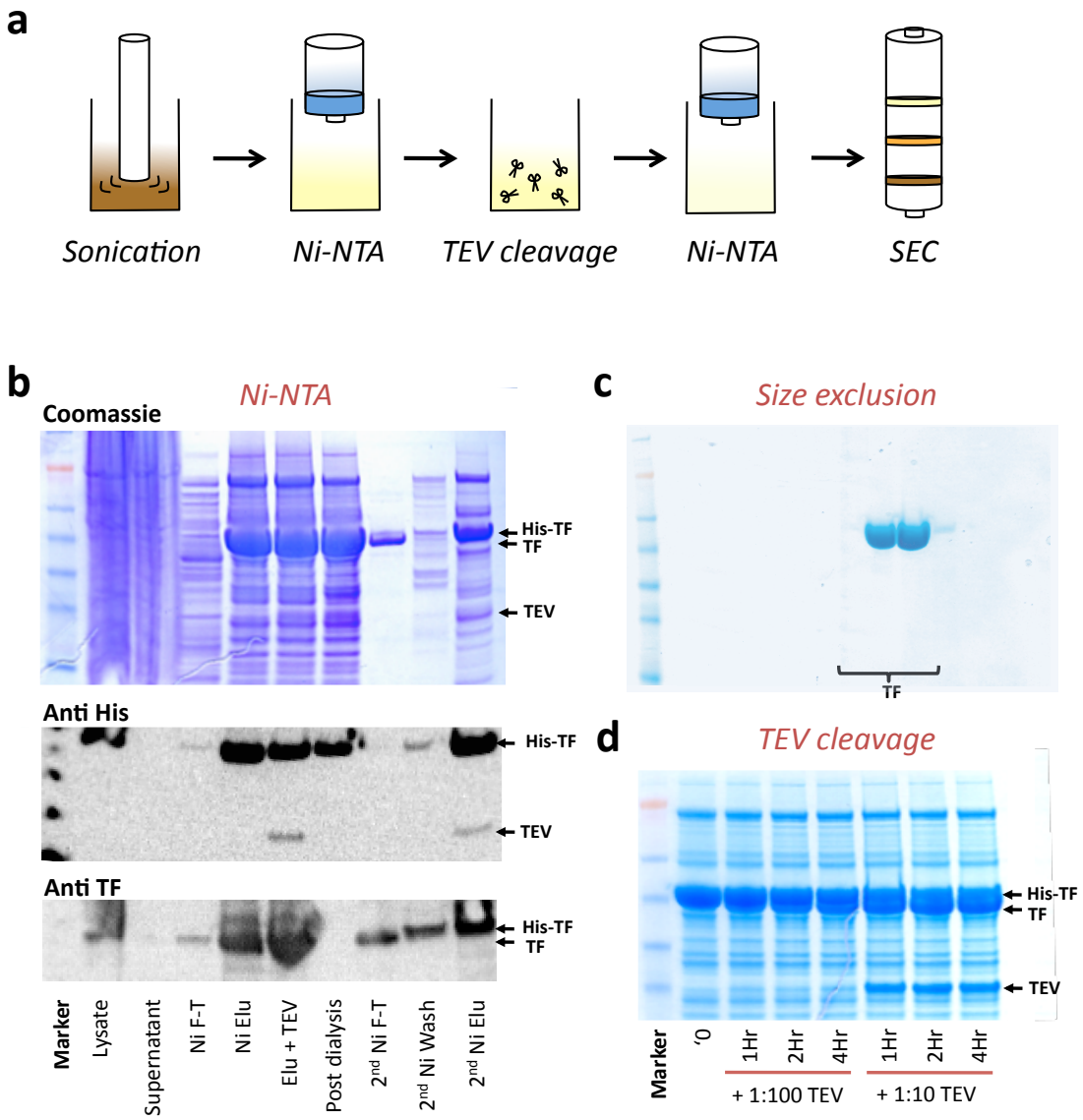

Figure 3.3: Trigger factor purification procedure. (**a**) Summary of the TF purification procedure. (**b**) SDS-PAGE stained by Coomassie stain and Anti-His and Anti-TF western blots of the first four purification steps. (**c**) SDS-PAGE stained by Coomassie stain of the final size exclusion chromatography step of the TF purification procedure. (**d**) SDS-PAGE of the optimisation of the TEV cleavage reaction.

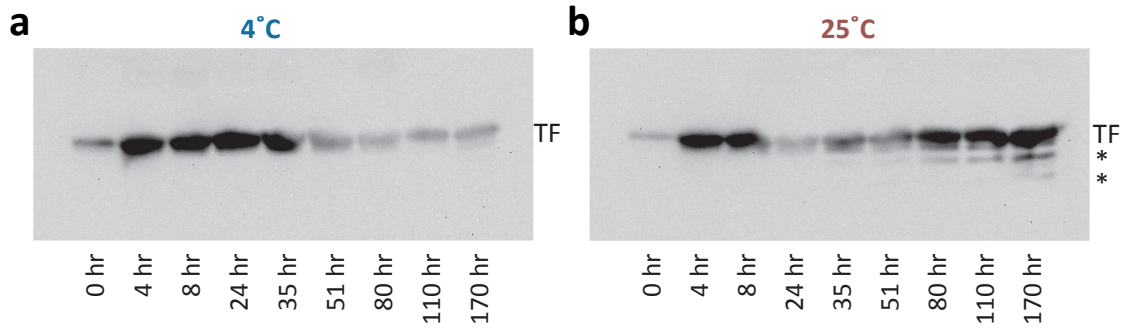

Figure 3.4: Characterisation of TF stability. Anti-TF detected western blots of TF stability timecourses at 4°C (**a**) and 25 °C (**b**). Degradation products are marked with stars.

TF stability : The stability of TF under conditions compatible with RNC investigations described further in Chapters 4 and 5 (in Tico buffer, pH 7.5, 4 , 10 or 25°C) was monitored over the duration of 7 days by Western Blot immuno-assay (Fig 3.4). At 4°C, TF was found to be stable up to 170 hours (7 days) where first traces of smaller degradation products were detected. Degradation of TF was more rapid at 25°C with a sample life-time around 50 hours.

3.3.3 Characterisation of TF^{Ile} by NMR spectroscopy

3.3.3.1 ^{1}H-^{13}C methyl TROSY HMQC of selectively protonated ^{1}H-^{13}C isoleucine residues in TF

The Transverse Relaxation Optimised Spectrometry (TROSY) ^{1}H-^{13}C HMQC experiment selects for slow relaxing, long lived components of the NMR signal of coupled spins while broadening out faster relaxing components. In large biomolecules, where signal is lost because of the slow tumbling speed, this method is highly advantageous to study the behaviour of methyl and amino groups on a perdeuterated background [199]. A more detailed description of TROSY and its advantages can be found in Sections 2.1.3 and 2.1.5.

^{1}H-^{13}C HMQC (Heterologous Multiple Quantum Correlation) spectra were recorded of uniformly ^{13}C labelled TF (100 μM in Tico, 99% D_2O pH 7.5) and selective isoleucine ^{1}H-^{13}C methyl labelled TF (from hereon referred to as TF^{ile}, 150 μM in Tico, 99% D_2O pH 7.5) at 25°C. The isoleucine side-chain methyl group region of the spectrum shown in Fig 3.5.a demonstrates the striking improvements achieved by the combination of TROSY-optimised experiments and the employed selective labelling strategy.

The H_6-TF^{ile} spectrum shows 20 well-resolved cross-peaks out of 25 expected isoleucine resonances. This is in contrast to the two highly broadened and overlapping peaks in the spectrum of uniformly ^{13}C-labelled TF, which correspond to the highest intensity peaks

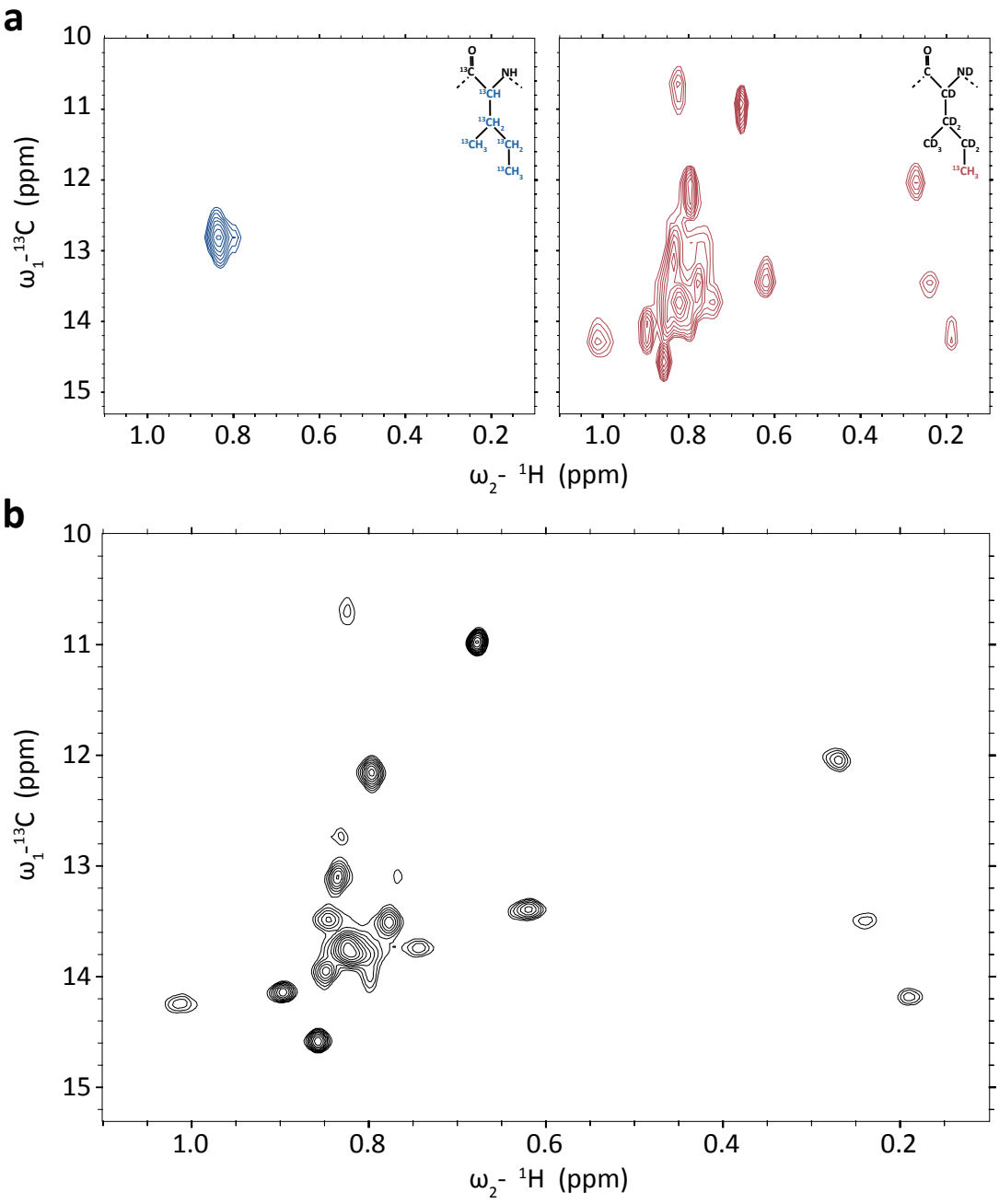

Figure 3.5: Selective protonation of TF isoleucine residues. (**a**) Isoleucine region of ^{1}H-^{13}C TROSY HMQC spectra of uniformly ^{13}C-labelled TF (blue) and selective ^{1}H-^{13}C labelled side-chain methyl groups of TF isoleucine residues (pink). (**b**) High resolution ^{1}H-^{13}C TROSY HMQC spectrum of H_6-TF^{ile}.

observed in the TF^{ile} spectrum. Integration of the total signal in both spectra recorded under similar experimental conditions (see Section 2.3) and normalised for concentration differences, shows that selective labelling of the isoleucine side-chain methyl groups results in a ∼13 fold increase of the NMR signal in the isoleucine region of the TF NMR spectrum (0-1ppm in ^{1}H dimension 10-15 ppm in ^{13}C dimension) relative to the uniformly ^{13}C labelled sample.

The H_6-TF^{ile} ^{1}H-^{13}C HMQC spectrum displays 10 well-dispersed peaks in the high (0-0.6 ppm) and low (0.9-1 ppm) field regions of the spectrum which provide excellent probes for the monitoring of the TF structure under varying experimental conditions. The observed spectral dispersion correlates with expected values for a well-folded protein. The resonances in the overlapped region between 0.6 and 0.9 ppm originate from residues in more flexible structural elements of the TF protein such as loops and unstructured regions. Protein perdeuteration, selective isotopic labelling and careful selection of NMR experiments have enabled the improvement of the TF NMR spectra to a extent where a detailed, high resolution characterisation of the large TF molecule is possible.

3.3.3.2 The effect of temperature on the quality of the TF^{Ile} HMQC spectra

^{1}H-^{13}C TROSY HMQC spectra of H_6-TF^{ile} were recorded across various temperatures (10, 15, 20, 25, 37°C, Fig 3.6.a) in order to determine ideal experimental conditions. Temperature significantly affects the rotational correlation time of molecules and thus impacts the transverse relaxation rates. This, in turn, affects linewidth as further described Section 2.1.2. ^{1}H-^{13}C TROSY HMQC spectra at lower temperatures are thus typically associated with broader resonance linewidths. Indeed, significant broader resonances of weaker intensity are observed for H_6-TF^{ile} as the temperature decreases (Fig 3.6.b).

The observed line-broadening is however not uniform across all resonances. Certain resonances shown significantly more intensity at 20°C than at 25°C (Fig 3.6.b) which is indicative of chemical exchange. The chemical exchange broadening is more pronounced at temperatures above 20°C indicating that most resonances are in slow exchange at low temperatures and gradually approach an intermediate exchange regime around 25°C as k_{ex} increases. The origin of the chemical exchange is likely to be found in the TF dimerisation reaction although structural re-arrangement between or within domains cannot be excluded. The observed chemical exchange is the topic of further discussion in Section 3.5.

The absence of notable exchange broadening at 20°C results in a slight improvement in the signal to noise relative to 25°C. However, since pre-existing data relating to the NMR of ribosome-nascent chain complexes (see Section 1.3.3) has been acquired at 25°C, most experiments through-out this thesis have been acquired at this temperature for consistency.

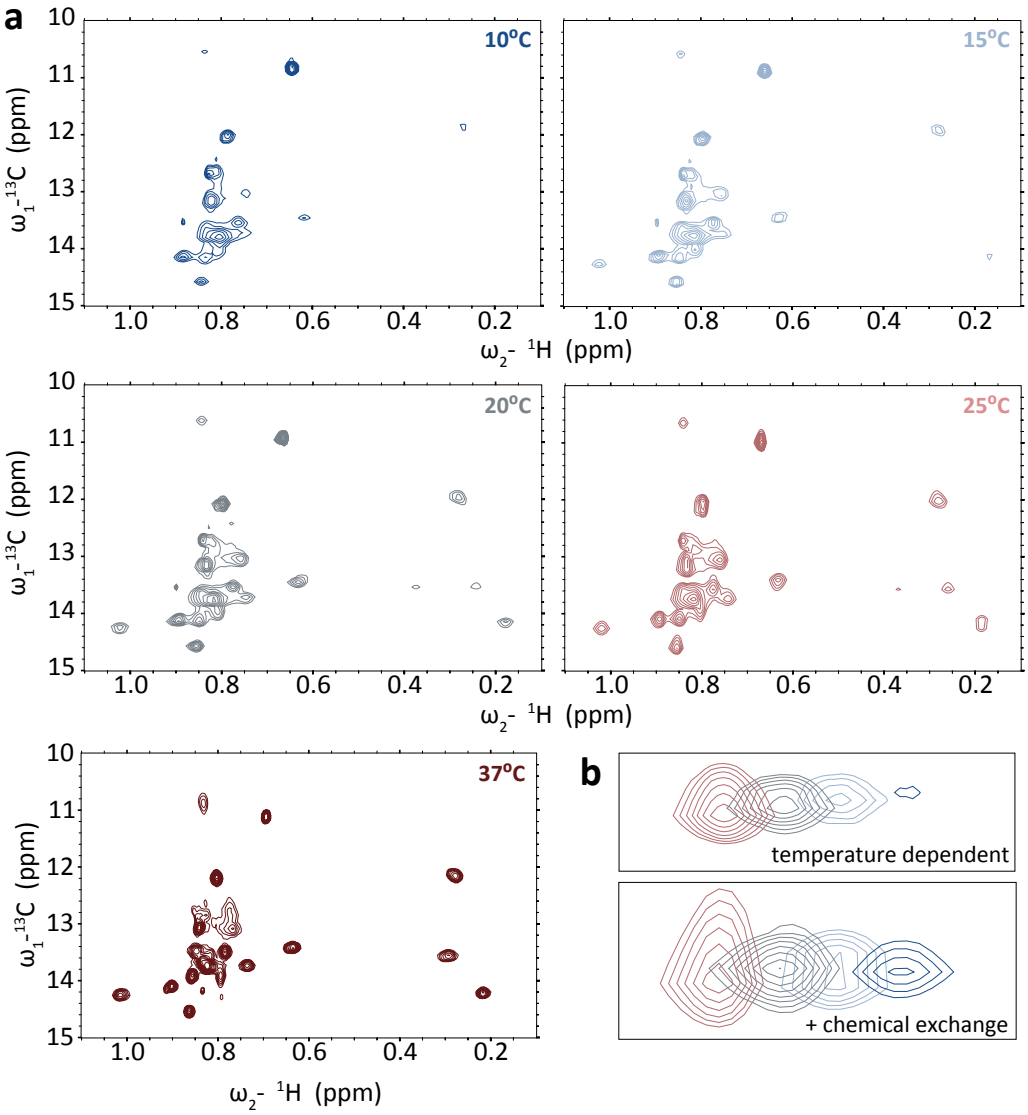

Figure 3.6: Trigger factor temperature titration. (**a**) ^{1}H-^{13}C TROSY HMQC spectra of H_6-TF^{Ile} at 10, 15, 20, 25 and 37°C. (**b**) Two examples of H_6-TF^{Ile} resonances across temperatures (10-25°C, colours as in **a**) illustrating the heterogeneity in temperature dependent broadening indicating the presence of chemical exchange.

While increasing the temperature of acquisition further to 37°C significantly increases the signal intensities, TF stability was found to be compromised with degradation product appearing as soon as 4 hours into acquisition.

3.3.4 Expansion of the selective labelling strategy to Ile, Leu and Val residues.

The selective ^{1}H-^{13}C isoleucine side chain methyl labelling strategy was extended to Cδ and Cγ methyls of leucine and valine residues respectively by the inclusion of the metabolic precursor to both these amino acids: isovalerate. Selectivity for Cδ/γ1 or Cδ/γ2 was not achieved and each valine and leucine residues thus appears in the TROSY HMQC spectra as two, potentially distinct, cross peaks.

The ^{1}H-^{13}C methyl-TROSY HMQC, Fig 3.7, was recorded of 100 μM TF ^{ILV} in Tico (pH 7.5) at 25°C. Using this labelling strategy, the number of available structural and dynamic probes in the TROSY HMQC spectrum was increased from 20 for Ile-labelling alone to 126. 98 distinct residues were labelled using this strategy giving rise to a maximum of 171 NMR resonances in the case where each leu/val residues corresponds to two distinct signals. Of the detected resonances, 70 (from 10) are sufficiently well-dispersed to be monitored across a range of experimental conditions.

The advantage of an increased number of probes is potentially offset by an increased proton density over the protein molecule which can lead to enhanced relaxation rates of the side-chain methyl groups and thus weaker signals through spin diffusion. Comparison of an equivalent TF^{ile} and TF^{ILV} labelled NMR data set, shows a reduction of the signal to approximately 85% due to increased spin diffusion in the protein while the number of structural probes has been quadrupled.

3.3.5 Assignment of ^{1}H-^{13}C TROSY HMQC resonances of TF^{ILV}.

A partial assignment of TF residues was provided by Prof Babis Kalodimos at Rutger University, NJ shown in Fig 3.7. 95% of Ile methyl resonances have been assigned using a 'divide and conquer' assignment strategy (Babis Kalodimos, personal communication 2013). Individual domains were assigned in isolation and the assignments transferred to the full-length protein spectrum. Ambiguities in the assignment were resolved by mutagenesis (typically x to ala) and comparison of detected NOEs from neighbouring, assigned residues with the known *E. coli* TF crystal structure (described in Section 1.2.2.1). The assignment of Leu and Val residues is only 32% complete through the significant overlap and difficulties to transpose domain assignments onto the full-length spectrum.

The assignment covers a significant part of TF; it is noticeably thin in the ribosome binding domain with only 7 of 27 Ile, Leu and Val residues assigned. This particular domain is the most problematic in the current assignment strategy since it appears to be adopting a non-native like dimer structure, whose spectrum does not bear semblance to the full-length one [130].

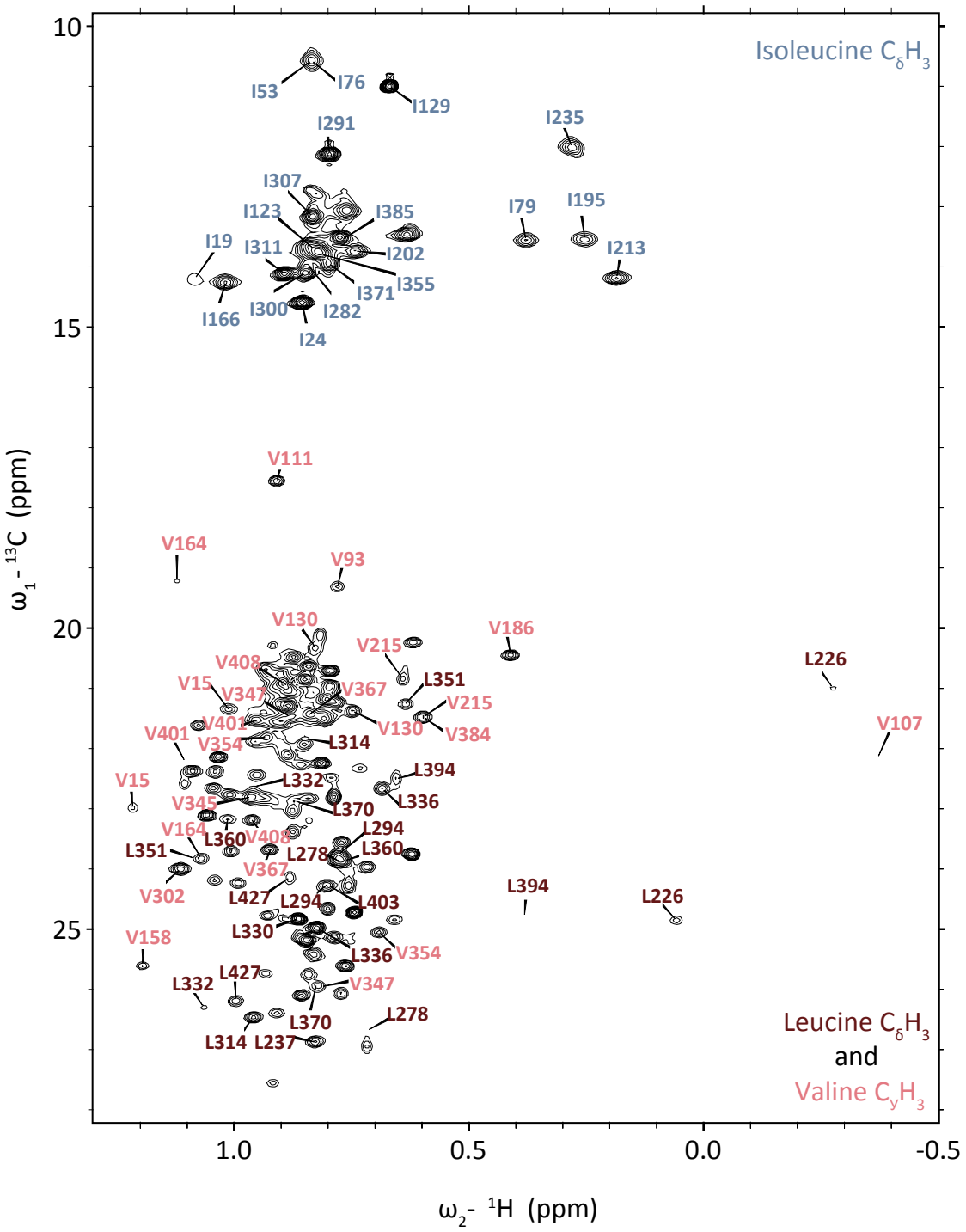

Figure 3.7: Trigger factor ILV assignment

3.4 Mathematical modelling of the five state TF equilibrium.

Within the cell, trigger factor is likely to exist in a dynamic multi-state equilibrium as depicted in Fig 1.10 and simplified in Fig 3.8. In the absence of substrates, TF adopts a dimeric form which is the topic of further investigation in Section 3.5.2. It is the monomeric state of TF that interacts with its various partners in the cytosol including inactive and actively translating ribosomes, folded and unfolded isolated proteins. This introduces competition between the different association reactions providing a potential means of directing TF activity where it is required.

The study of one particular TF state is complicated by the need to account for the competing reactions. These competing equilibria render the study of TF-substrate interactions non trivial. In particular in the study of the TF interaction with an RNC, both dimerisation and unoccupied ribosomes compete with the RNC for TF. Even with the highly pure and high occupancy RNC complexes that can be produced, eliminating the fifth competing isolated protein equilibrium, four different TF states; dimer, monomer, 70S-bound and RNC-bound, need to be accounted for. In this Section (3.4), an overview of our current knowledge on the TF equilibria is presented and a mathematical model is built to aid the deconvolution of competing reactions in further experiments presented in this thesis.

3.4.1 An analysis of the TF equilibrium - background.

The distribution of TF over the various states of its dynamic equilibrium is dependent on the availability of substrates as shown in Fig 3.8. The isolated association reactions have been extensively studied and a comprehensive overview of the published dissociation constants are presented in Tables 3.1-3.3. Equilibrium distributions and kinetics have been studied across a wide range of temperatures (4-30°C). Apparent differences across temperature are however smaller than between different studies under similar experimental conditions and thus the overview presented here does not permit to draw conclusions about the temperature dependence of the various reactions.

The methods used to investigated the equilibrium behaviour of the four TF reactions described in Fig 3.8 vary across the different studies with overall prevalence of fluorescence and analytical ultr-centrifugation (AUC). The values obtained for each TF equilibrium are however within the same order of magnitude, with a few notable exceptions highlighted below.

Both the TF dimerisation and ribosome binding reactions have been found to occur with affinities ranging from 1 to 2 μM (Table 3.10) [148, 160] while the TF affinity for

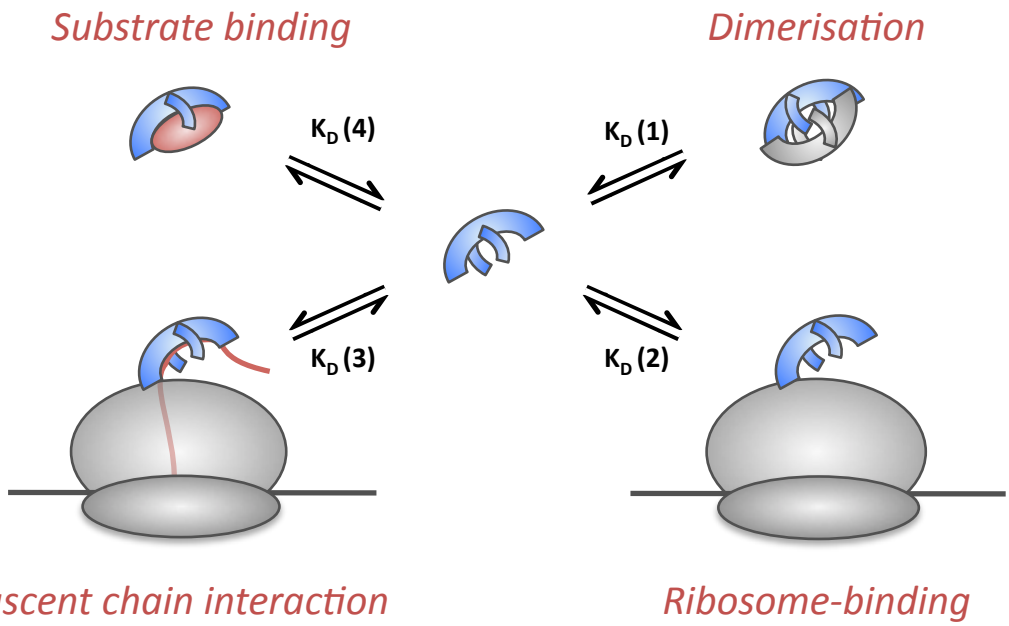

Figure 3.8: The dynamic trigger factor multi-state equilibrium. TF (blue) exist in a dynamic equilibrium between various substrate (red) bound states. The monomer self-associates into a dimeric form. TF transiently interacts with inactive ribosomes while preferring nascent chain occupied ribosomes. TF further interacts with low affinity with some unfolded and folded protein substrates independently from the ribosome.

RNCs is consistently higher with values around 100 nM with variations up to 690 nM and down to 2 nM dependent on the specific nascent chain present (Table 3.11) [148, 240].

Three out of four values for the TF dimerisation reaction are in close agreement (Table 3.10) with an average value of Kd_1 (as in Fig 3.8) of 1.25 μM. More variation is observed in the measured values of the dissociation constant for the interaction of TF with inactive ribosomes ranging from 300 nM to 2.1 μM. An average value of 1.1 μM puts the dissociation constant for ribosome binding just below that of TF dimerisation.

As described, values for the dissociation constant of the TF interaction with RNCs cover a larger range from 2 nM up to 690 nM (Table 3.11). All obtained values are significantly lower than the dissociation constant for the TF interaction with inactive ribosomes. The affinity of TF for RNC substrates presents both sequence and length dependence with a general preference for nascent chain substrates between 100-200 residues in length. A recent study [162], has challenged some of results presented in preceding studies into the kinetics of TF association with RNC substrates. It found significantly faster equilibrium kinetics with half-lives <0.1 ms as opposed to 10 s as previously determined and lower values for the dissociation constants of the interaction of TF with both ribosomes and RNC substrates ($\sim$100 nM and $\sim$ 2 nM respectively) (Table 3.11). The discrepancies with pre-

viously published results were attributed to the choice of fluorescent label (BADAN) that is proposed to induce a second, rate-limiting, conformational change in TF before 70S association. The rapid kinetics observed in this study might also affect the interpretation of the co-sedimentation studies which rely on a long lived interaction for pull-down of the ribosome-bound TF fraction. It is thus likely that the apparent K_{d} values presented in the co-sedimentation study present overestimations of the dissociation of the TF interaction with ribosomes and RNC complexes.

A further caveat in some of the literature K_{d} values described above is that the acquired data have been fitted to a single reaction, independently of other equilibria present, notably the TF dimerisation reaction. In the fluorescence based studies, protein concentrations used to experimentally determine the dissociation constants are in the high nM range,

Table 3.1: Literature overview of the kinetics of trigger factor dimerisation and binding to ribosomes

$K_{\mathbf{d}}$ (nM)	$\mathbf{k}_{on}$ ($\mathrm{mM^{-1}s^{-1}}$)	$\mathbf{k}_{off}$ ($\mathrm{s^{-1}}$)	**Temp** (°C)	**Method**	**Publication**
Dimerisation					
1000*		-	-	Size Exclusion chromatography	[159]
18000*		-	20	AUC	
1800*		-	20	AUC	[160]
900*		1	30	fluorescence anisotropy	[148]
Ribosome interaction					
300*	-	-	4	gradient centrifugation	[145]
1200*	-	-	30	co-sedimentation	[159]
1100*	-	0.036	20	fluorescence anisotropy	[160]
800*	-	-	30	fluorescence anisotropy	[148]
1100 ± 300	-	-	4	co-sedimentation	[240]
2100	21 ± 9	0.045 ± 0.001	30	fluorescence	[161]
140	0.085 ± 0.005	12 ± 2	25	fluorescence	[162]
70 ± 10	-	-	25	FRET	

* no errors available

Table 3.2: Literature overview of the kinetics of trigger factor binding to RNCs

RNC	$K_\mathbf{d}$ nM	$\mathbf{k}_{on}$ $mM^{-1}s^{-1}$	$\mathbf{k}_{off}$ s^{-1}	**Temp** °C	**Method**	**Publication**
FL*	30	-	-	30	fluorescence	[148]
RpoB-60	690 ± 77	-	-	4	co-sedimentation	[240]
RpoB-80	141 ± 13	-	-			
RpoB-100	71 ± 7	-	-			
RpoB-133	53 ± 9	-	-			
Bop-81	670 ± 194	-	-			
Bop-98	316 ± 66	-	-			
Bop-136	422 ± 86	-	-			
Lep-100	311 ± 113	-	-			
ICDH-27	496	96 ± 1	0.047 ± 0.002	20	fluorescence	[161]
ICDH-66	342	142 ± 19	0.049 ± 0.001			
ICDH-108	269	186 ± 8	0.048 ± 0.001			
ICDH-177	393	103 ± 12	0.037 ± 0.001			
ICDH-318	341	86 ± 2	0.029 ± 0.002			
RpoB-23	480	95 ± 6	0.046 ± 0.001			
RpoB-36	276	163 ± 2	0.045 ± 0.001			
RpoB-100	103	169 ± 6	0.017 ± 0.001			
RpoB-148	224	116 ± 4	0.026 ± 0.001			
RpoB-190	89	146 ± 8	0.013 ± 0.001			
RpoB-233	149	137 ± 10	0.020 ± 0.001			
RpoB-325	213	78 ± 4	0.017 ± 0.001			
RpoB-516	292	71 ± 6	0.021 ± 0.001			
MetK-30	453	99 ± 4	0.045 ± 0.002			
MetK-60	308	141 ± 4	0.043 ± 0.002			
MetK-101	138	168 ± 9	0.023 ± 0.002			
MetK-186	279	109 ± 8	0.030 ± 0.001			
FL-31	351	140 ± 9	0.049 ± 0.001			
FL-106	193	167 ± 5	0.032 ± 0.001			
FL-223	181	151 ± 4	0.027 ± 0.002			
SH3-wt	405	108 ± 4	0.044 ± 0.001			
SH3-m10	314	149 ± 4	0.047 ± 0.001			
Barnase	440	99 ± 5	0.044 ± 0.001			
Barnase-95	181	154 ± 5	0.028 ± 0.001			
proOompA75	2	0.2 ± 0.02	0.4 ± 0.1	25	fluorescence	[162]
	2.5 ± 0.5	-	-		FRET	
HemK75	2.5	0.24 ± 0.02	0.6 ± 0.2		fluorescence	
	10 ± 2	-	-		FRET	
Lep75	110	0.11 ± 0.015	12 ± 2		fluorescence	
	90 ± 10	-	-		FRET	

* no errors available

Table 3.3: Literature overview of the kinetics of trigger factor binding to isolated proteins

Substrate	$K_\mathbf{d}$ μM	$\mathbf{k}_{on}$ $mM^{-1}s^{-1}$	$\mathbf{k}_{off}$ s^{-1}	**Temp** °C	**Method**	**Publication**
RCM-La*	1.7	-	-	15	fluorescence	[241]
RCM-T_1*	11.5	-	-			
apa-La*	7.8	-	-			
S7*	3.5	5.03	0.0175	25	biacore	[84]
PhoA^{a*}	25	0.3	1000	22	ITC	[130]
PhoA^{b*}	200	-	-			
PhoA^{c1*}	150	-	-			
PhoA^{c2*}	170	-	-			
PhoA^{c1-c2*}	50	-	-			
PhoA^{c1-c3*}	10	-	-			
PhoA^{a-c*}	2	1.45	50			
PhoA^{d*}	200	-	-			
PhoA^{e*}	18	-	-			
PhoA^{d-e*}	4	-	-			
PhoA^{f*}	200	-	-			
PhoA^{g*}	50	-	-			
PhoA^{f-g*}	14	-	-			

* no errors available

below the K_d of dimerisation. It is thus reasonable to assume in these cases that dimer dissociation does not play a limiting role in the substrate-binding reactions and does not significantly affect the measured dissociation constants. At higher concentrations however, competition by the dimerisation reaction becomes significant and the published apparent K_d values are thus likely to present an overestimation of K_d values of the TF monomer associating with either ribosomes or RNC (reactions 2 and 3 in Fig 3.8).

The study by Martinez-Hackert *et al.* [84] focussed on the unique finding that TF interacts with folded protein substrates independently of the ribosome. Since the exact mechanism of binding is unknown, the distinction between dimeric TF or monomeric TF interacting with the substrate cannot be made. The available literature on isolated substrate interactions is limited, summarised in Table 3.3. It should be noted that the current understanding of the TF interaction with isolated substrate proteins is highly biased towards tightly interacting substrates for their initial identification and thus provides an overall overestimation of the likely 'average' affinity of TF for this type of isolated protein substrate.

3.4.2 Mathematical model of TF equilibria

The complex TF equilibrium as shown in Figure 3.8 can be described mathematically by the following expressions:

Dimerisation:

$$Kd_1 = \frac{[Monomer]^2}{[Dimer]} \tag{3.4.1}$$

Ribosome-binding:

$$Kd_2 = \frac{[Monomer][70S]}{[TF{:}70S]} \tag{3.4.2}$$

RNC-binding:

$$Kd_3 = \frac{[Monomer][RNC]}{[TF{:}RNC]} \tag{3.4.3}$$

Isolated protein binding:

$$Kd_4 = \frac{[Monomer][Substrate]}{[TF{:}Substrate]} \tag{3.4.4}$$

These expressions are however a function of concentrations that cannot be directly measured; n.b free and complexed TF. Inclusion of the definition of these concentrations as a function of the original protein concentrations in the experiments as below allows us to further define the system.

$$[TF]_0 = 2[Dimer] + [Monomer] + [TF{:}70S] + [TF{:}RNC] + [TF{:}Substrate] \tag{3.4.5}$$

$$[Substrate]_0 = [Substrate] + [TF{:}Substrate] \tag{3.4.6}$$

A final set of equations required to relate the RNC to the ribosome concentrations can be defined by introducing the nascent chain occupancy of the ribosomes α as follows:

$$(1-\alpha)[Ribosome]_0 = [70S] + [TF{:}70S] \tag{3.4.7}$$

$$\alpha[Ribosome]_0 = [RNC] + [TF{:}RNC] \tag{3.4.8}$$

This system of 8 simultaneous equations is sufficiently restrained to numerically solve for the TF population distribution when experimental parameters such as the starting 70S and TF concentrations, the 70S nascent chain occupancy and all four dissociation constants are known.

The mathematical description of the simplified TF equilibrium (Fig 3.8) was numerically solved using mathematical software package Mathematica. Fig 3.9 presents a sample input script as used for the prediction of the TF population distribution across the various substrate-bound states under specific experimental conditions throughout this thesis. In the remainder of this Section, some observations are made about the general behaviour of the TF population within the multi-state equilibrium. The values of the dissociation constants describing the various reaction used here are: 1.25 μM for the dimerisation reaction (Kd_1), 1 μM and 100 nM for the TF interaction with inactive ribosomes (Kd_2) and 10 nM and 100 nM for the interaction of TF with the ribosome-nascent chain complexes (Kd_3, Fig 3.8).

3.4.3 The TF monomer-dimer equilibrium

The simplest reaction described by the model introduced in Section 3.4.2 is the dimerisation of TF in the absence of heterologous substrates. Values described for the dissociation constant of this reaction are 0.9 μM [160], 1.8 μM [148] and 1 μM [159]; an average value of 1.25μM will be used in the predictions made throughout the remainder of this chapter.

As is apparent from Fig 3.10.a, at 1.25 μM, the K_d value of the reaction, the TF population is equally distributed over monomer and dimer species. At lower concentrations ($<K_\mathrm{d}$) the TF monomer is favoured while at higher concentrations the dimer forms the major species.

NMR investigation of large macromolecules such as TF using the selective labelling strategies outlined in Section 3.3 are feasible from concentrations of 10 μM upwards. For detailed, high resolution study of macromolecules, 100 μM concentrations are a minimum requirement (for more details, refer to Section 2.1.5).

While the monomer and dimer population fractions are equal at 1.25 μM, this differs from the concentration at which both monomer and dimer concentrations are the same. This is in contrast to typical heterologous complexes studied where the complex concentration equals that of the reactants at the K_d because the TF dimer is made up of two TF monomers. The monomer and dimer concentrations coincide at 3.8 μM. This distinction is important since both TF molecules in the TF dimer contribute to the dimer NMR signal. In the case where the TF dimer is symmetrical, the NMR signal from a specific residue in one monomer of the dimer will be very likely to overlap with that of the second monomer because their chemical environments will be closely similar. As a consequence,

```
ɣ[T0_, R0_, S0_, α_, K1_, K2_, K3_, K4_] :=
  Module[{ɣD, ɣM, ɣR, ɣN, ɣS, eqns, sol, d, a, b, c, n, r, m, s},
    eqns = {
      K1 == m^2/d,
      K2 == m r/a,
      K3 == m n/b,
      K4 == m s/c,
      T0 == m + 2 d + a + b + c,
      (1 - α) R0 == r + a,
      α R0 == n + b,
      S0 == c + s
     };
    (*NSolve[eqns/.T0→2*^-6/.K1→1*^-6/.K2→1*^-6/.K3→1*^-7/.R0→10*^-6/.α→.9/.S0→10*^-6/.
      K4→1*^-6]*)
    sol = Select[
       NSolve[eqns]
       , (Re[m /. #] > 0 && Re[n /. #] > 0 && Re[a /. #] > 0 && Re[b /. #] > 0 &&
          Re[c /. #] > 0 && Re[s /. #] > 0 && Re[r /. #] > 0 && Re[d /. #] > 0) &][[1]];
    (*Print[(m+2d+a+b+s)/.sol];*)
    ɣD = 2 d/T0 /. sol;
    ɣM = m/T0 /. sol;
    ɣN = b/T0 /. sol;
    ɣR = a/T0 /. sol;
    ɣS = c/T0 /. sol;

    {ɣD, ɣM, ɣR, ɣN, ɣS}
   ];
```

Figure 3.9: Mathematical description of the TF multi state equilibrium as Mathematica input.

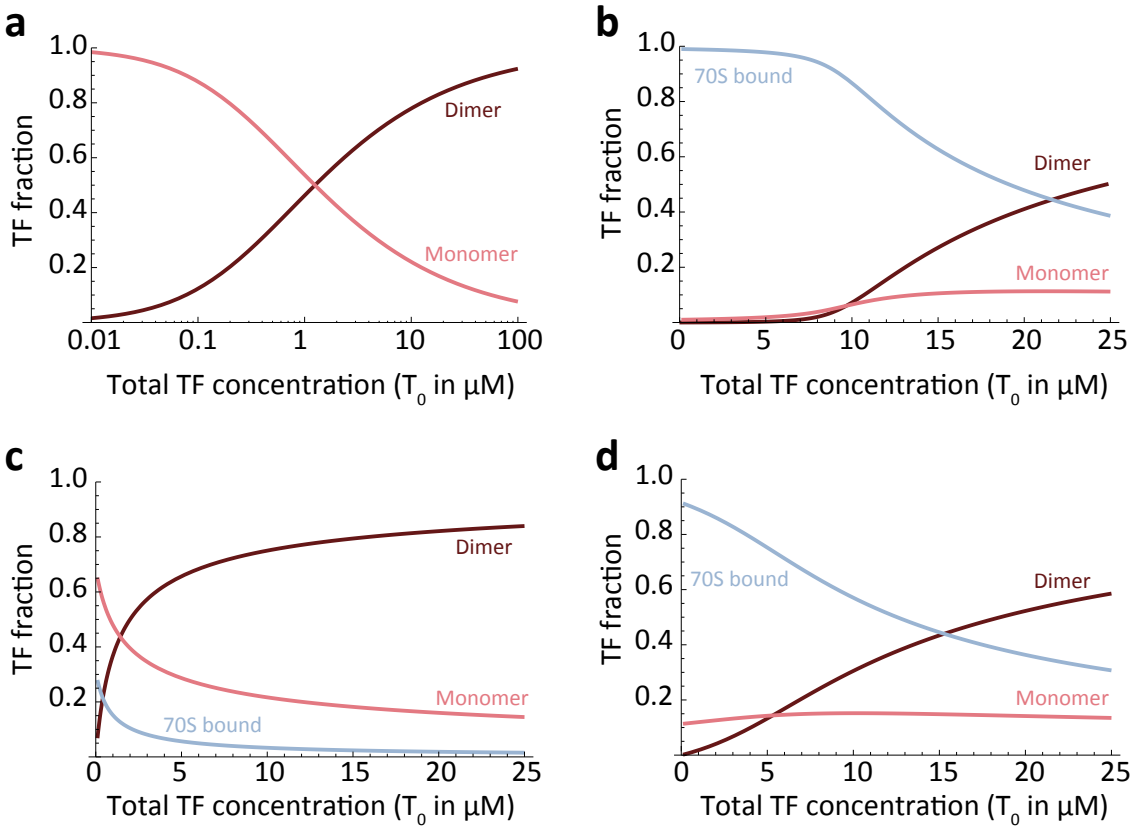

Figure 3.10: Mathematical description of the TF dimerisation and 70S association equilibria. (**a**) TF population distribution over monomer (pink) and dimer (red) states with increasing TF concentration (on a logarithmic scale). (**b**) TF population distribution over monomeric, dimeric and 70S-bound (blue) states with increasing TF concentration in the presence of 10 μM 70S and a predicted affinity for the ribosome of 100 nM. (**c**) TF population distribution over monomeric, dimeric and 70S-bound states with increasing TF concentration in the presence of 0.5 μM 70S and a predicted affinity for the ribosome of 1 μM. (**d**) TF population distribution over monomeric, dimeric and 70S-bound states with increasing TF concentration in the presence of 10 μM 70S and a predicted affinity for the ribosome of 1 μM.

NMR signal integrals of the resonances originating from the same residue in the monomer and dimer will be equal at 1.25 μM.

However, in the case where the TF dimer interface is asymmetrical, the same residue might experience a different chemical environment from one sub-unit of the dimer to another, distinct again from the same residue in the monomer population. This would give rise to two distinct NMR signals for the same residue in the TF dimer. Since the signals no longer add up, the signal integrals will be equal to that of the monomer signal at 3.8 μM where the monomer and dimer concentrations are the same. This phenomenon is described further in Section 3.5.1.1.

3.4.4 Competition between TF dimerisation and ribosome binding

Addition of ribosomes to TF introduces competition between two equilibria of similar affinities. While the average published value for dimerisation is 1.25 μM, ribosome binding was found to occur with an average K_{d} of 1.1 μM [148, 160]. Since ribosome binding is slightly favoured, the gradual introduction of higher concentrations of 70S to the model, results in a shift of the dimerisation equilibrium (note the intersect of the red and pink lines in Fig 3.10.b) to the right towards higher TF concentrations. For instance, 0.5 μM ribosomes (Fig 3.10.c) shifts the point of equal populations of monomer and dimer towards 1.4 μM while 10 μM 70S displaces the equilibrium to 4.9 μM. As a result, the monomer fraction can be favoured over the dimer fraction at TF concentrations up to 6 μM. It should be noted that although the relative monomer fraction can be manipulated at will within a concentration range of 1-10 μM TF by varying the added concentration of 70S, the absolute monomer concentration never exceeds that in the absence of 70S.

The changes in the relative sizes of the monomer and dimer populations upon addition of ribosomes are relevant to the NMR investigation that follows since NMR spectra display population weighted average signals. NMR spectra of TF in presence and absence of ribosomes cannot be readily compared as the ribosome addition results not only in the contribution of the 70S-bound state to the spectrum but also in a shift in the monomer/dimer equilibrium. A good understanding of the interplay between the dimerisation reaction and ribosome-binding could thus aid in the deconvolution of the various contribution to the NMR spectrum (see Section 1.5.1.3).

Remarkably, at ribosome concentrations above 1.1 μM, i.e the ribosome binding dissociation constant K_{d3}, the monomer fraction reaches a 'steady state' whose relative size remains constant over a large range of TF concentrations. The direct exchange of 70S-bound TF to dimer is observed instead (Fig 3.10). This observation suggests that the monomer contribution could potentially be eliminated in studies investigating the TF-ribosome interaction since the relative contribution to the NMR spectrum is concentration independent.

A tighter affinity of TF for the ribosome as indicated by [162] with a K_{d} of 100 nM abolishes the competition between the dimerisation reaction and ribosome-binding. Ribosome-binding is under these circumstances highly favoured and both TF dimer and monomer species do not accumulate significantly until saturation is achieved (Fig 3.10.b). The presence of a single TF species in solution (the ribosome-bound state) would simplify the interpretation of NMR results as further discussed in Section 4.5.

3.4.5 RNC occupancy is the determining factor in TF population distribution.

TF interacts with nascent chain-occupied ribosomes with an approximate ten-fold greater affinity than with inactive ribosomes. RNCs thus outcompete the other TF equilibria and the amount of TF bound to RNCs is directly correlated to the amount of RNC available as determined by the ribosome occupancy (Fig 3.11). For instance, for 10 μM TF and 10 μM 70S, at 90% occupancy, 80% of TF is RNC bound while for an occupancy of 30% this value is 28%. In fact, the curves in Fig 3.11 show that with increasing TF concentration, most TF is initially RNC bound, in the form of a plateau, until near saturation of RNC is reached at which point the relative fraction start decreasing.

Based on the reasonable assumption that the affinity of TF for nascent chain occupied ribosomes will always be equal or greater than that of inactive ribosomes, the exact K_{d3} value for the TF RNC interaction plays only a minor role in distributing the TF population over the various substrate bound states (Fig 3.11b).

The interplay of the various equilibria thus strongly favours RNCs over other substrates correlating with the idea that RNCs are TF's physiological substrates.

3.4.6 The TF equilibrium *in vivo*.

A number of predictions about the cellular distribution of TF can be made in a similar fashion although it is important to emphasise that the five state equilibrium as presented here is a simplification of the situation in the cell where TF can associate with a multitude of RNCs and isolated protein substrates, all with their respective dissociation constants. Assuming a normal distribution of the substrate affinities for TF in the cell however, an estimate can be made using average K_{d} values which would in principal not deviate far from reality.

Cellular concentrations of TF and the ribosome are approximately 50 μM and 20 μM respectively [242]. The extent of ribosome occupancy depends on the growth phase of the cell but it is estimated that during exponential growth ∼80% of ribosomes are active in translation in *E.coli* [243] .

Not taking into account any interactions with isolated protein substrates, the TF population distribution in the cell would be ∼54% dimeric, ∼37.5% ribosome-associated of which ∼85% is occupied by a nascent chain. This would imply that at any given time, 95% of all ribosomes are bound by TF. The remaining ∼8.5% TF exists in a monomeric form or 4.25 μM total per cell. The combination of a relative excess of TF and a relatively high affinity for the 70S could thus ensure that close to all ribosomes are occupied by TF at any given time, making TF an integral part of the protein biogenesis machinery. The two fold excess of TF over 70S also ensures that sufficient TF is present to permit

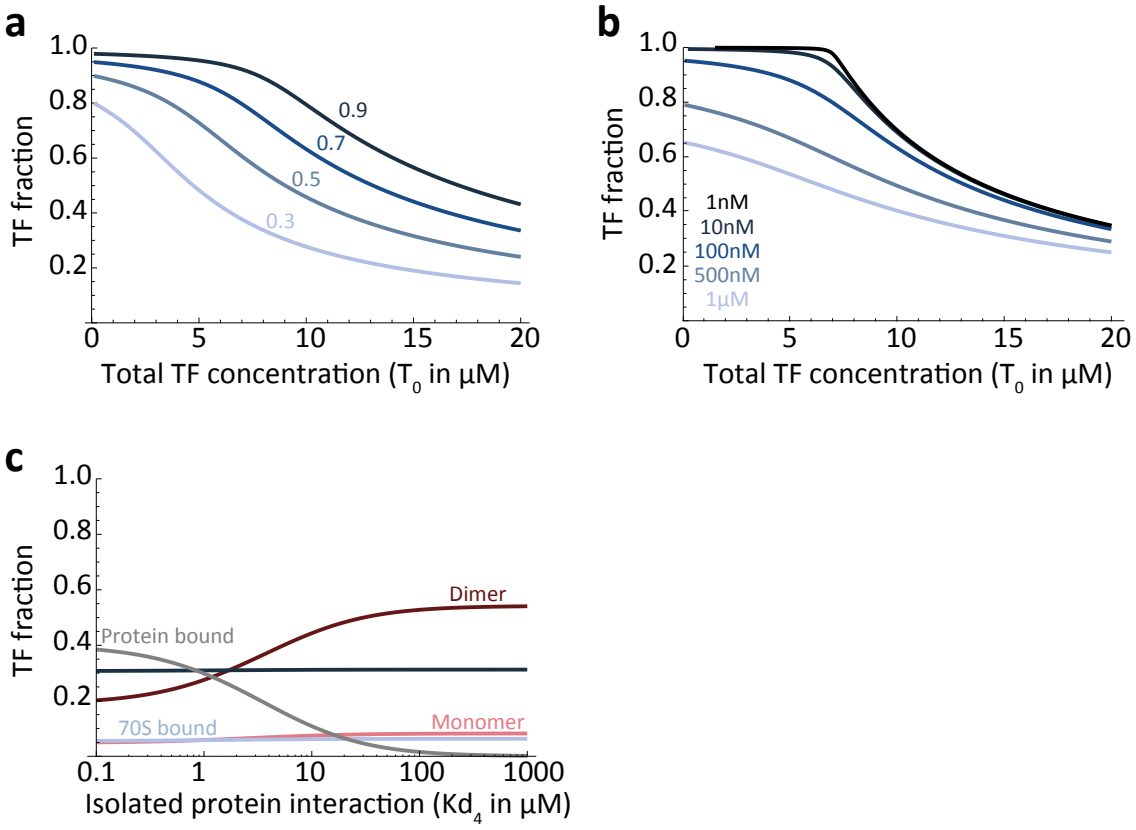

Figure 3.11: Mathematical description of the TF RNC association equilibrium and *in vivo* behaviour. (**a**) Behaviour of the RNC-bound TF population with increasing TF concentration for various RNC occupancy values (10μM ribosomes, NC occupancies from light to dark blue: 0.3, 0.5, 0.7, 0.9). (**b**) Behaviour of the RNC-bound TF population with increasing TF concentration for various dissociation constants for RNC association (from light to dark blue: 1 μM, 500 nM, 100 nM, 10 nM, 1 nM). (**c**) Predicted TF population distribution *in vivo* for varying affinities of TF for isolated protein substrates. Dimer in red, monomer in pink, 70S-bound in light blue, RNC-bound in dark blue and protein-bound in grey.

more specialist functions [86] such as the protection of fully synthesised, but uncomplexed ribosomal proteins (e.g S7) which interact with high affinity but exist in small populations, without TF interacting with other, more abundant proteins of lower affinity. Indeed the competition with the dimerisation reaction ensures that any reaction with a dissociation constant above 1.25 μM does not occur except in extreme cases of abnormal accumulation of a particular substrate in the cell.

The TF population estimates above are robust to deviations in the RNC affinities for TF. For example, a shift in RNC-binding K_{d} from 100 nM to 1 μM would result in only small changes in the population distribution, notably 54% to 59% dimeric and 31% to 26% RNC bound. Inclusion of isolated protein interactions into the equation results in a conversion of the dimeric form to a substrate bound state, the extent of which depends on

the concentration of substrate present and the TF affinity for the substrate (Fig 3.11.c). This conversion does not affect the fraction of RNC-bound TF at realistic estimates of the extremes in concentration (20 μM) and affinity (500 nM) of isolated protein substrates.

In summary, the five state TF equilibrium has been expressed mathematically and has been evaluated using literature values for the dissociation reactions. This model allows us to make a number of predictions on the behaviour of the TF population under specific experimental conditions; a finding which will be highly valuable in the interpretation of the recorded TF NMR data where the members of each state will contribute to various degrees to the observed spectrum.

3.5 NMR investigation of the TF dimerisation reaction

In the absence of heterologous substrates, TF exists in a dynamic equilibrium between a monomeric and dimeric state [148, 159, 160]. Although most interactions with heterologous substrates occur with monomeric TF [84, 130, 131, 159] the TF dimerisation reaction could play an essential role in regulating the interaction of TF with isolated protein substrates by competing with low affinity contacts and to act as storage form of excess TF in the absence of high affinity substrates. Indeed, the high protein content of the cell and the promiscuity of the TF interaction with hydrophobic substrates could exhaust the cellular pool of TF available for interaction at the ribosome surface in the absence of a competing mechanism or prevent native protein activity through transient interactions. The ongoing competition between substrate interaction and dimerisation might further form the basis of a release mechanism of TF from folding-competent nascent chain client proteins at the ribosome surface.

As discussed in Section 3.4, the dimeric TF complex was found to persist with a half-life ($t_{1/2}$) of $\sim$1 s (i.e the dissociation rate k_{off} $\sim$1 s^{-1}) and assembles with a dissociation constant of 1-2 μM [148, 159, 160]. The detailed study of proteins by NMR typically requires concentrations >100 μM. Under such experimental conditions, dimeric TF would be almost exclusively present in the absence of preferred substrates such as the ribosome (see Section 3.4 for more details on TF population distributions under various experimental conditions). Since the ongoing competition with the dimerisation reaction cannot be prevented in the study of the full-length TF molecule, it is important to account for its contribution towards the NMR spectrum even when the focus of the study is on the interaction of TF with a specific substrate. The remainder of this chapter therefore aims to characterise the NMR behaviour of the TF dimerisation reaction in order to improve our current understanding of this TF equilibrium and to understand the NMR spectral changes associated with the monomer/dimer equilibrium. This is particularly important for subsequent studies of substrate interactions investigated in Chapters 4 and 5.

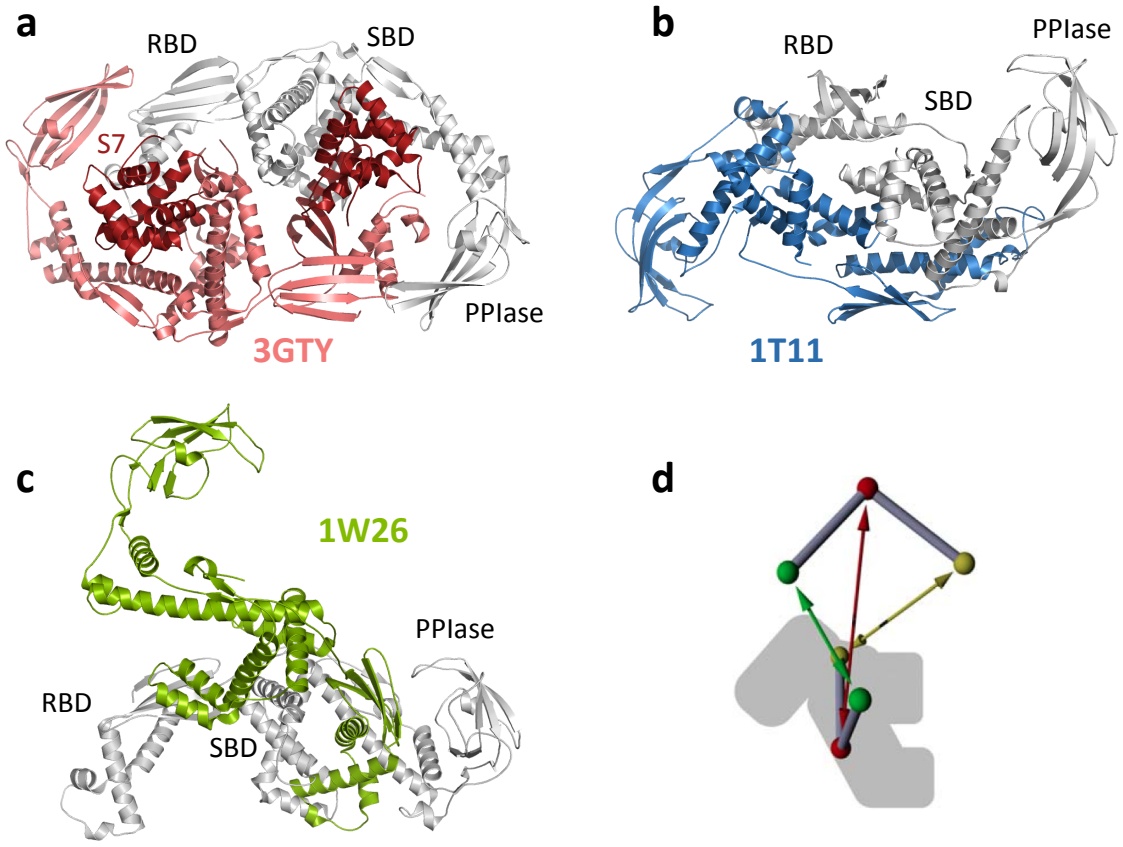

Figure 3.12: Overview of structures of dimeric TF (**a**) Structure of the *T.maritima* TF dimer in complex with the S7 ribosomal protein. (pdb 3GTY [84]). (**b**) Structure of the *V.cholerae* TF dimer (pdb 1T11 [238]). (**c**) Structure of the *E.coli* TF, dimeric structure from crystal symmetry (pdb 1W26 [126]). (**d**) Proposed orientation of the TF dimer in solution based on FRET distance restraints [148].

Our current structural understanding of the TF dimer presents a conflicting view of the dimerisation interface and, on a larger scale, the relative orientations of the two monomers as different X-ray crystal structures of TF dimers present a mix of symmetrical and non-symmetrical orientations (Fig 3.12). The *V.cholerae* TF dimer structure (Fig 3.12.b) presents a parallel, side-to-side packing of the monomers [238]. This truncated TF construct (44 residues missing from the C-terminus) was in later studies, however, shown to result in a non-native collapsed state [143]. The *T.maritima* TF complex with S7 (Fig 3.12.a) was shown to oligomerise to a 2:2 stoichiometry [84]. The two monomer 'cradle' regions, shown to interact with protein and nascent chain substrates [130–132] are buried within this complex. However, analysis of the structure indicates that the S7 substrate contributes significantly to the dimerisation interface interacting with large portions of the most hydrophobic regions of TF, intuitively buried in the native TF dimer interface. The

interaction between TF molecules is primarily mediated by the RBD and PPIase domains and the extremities of the 'arm' regions of the SBD. In the full-length *E. coli* TF crystal structure [126], the dimerisation interface is non-symmetric and the relative monomer orientation is perpendicular. The interaction is mediated by the RBD and SBD domains of one monomer with the SBD of the other and efficiently buries hydrophobic regions in one TF monomer only. Finally, distance restraints derived from inter-molecular FRET measurement propose a perpendicular orientation of the two TF monomers, ranking the distances from largest to smallest as follows: 14-14 >150-150 >326-326 [148].

The absence of a clear understanding of the TF dimerisation interface means that certain changes in the NMR signal upon substrate interactions cannot be readily attributed to dissociation of the dimer. We thus set out, in the study that follows, to build a model of the solution structure of the TF dimer using a range of NMR techniques. The size of the TF dimer (98 kDa) means that typical structure determination approaches based on NOE measurements cannot be readily undertaken. The strategy employed below thus relies on the determination of the dimerisation interface from basic two-dimensional fingerprinting spectra and medium range distance constraints from paramagnetic relaxation enhancement (PRE) measurements.

3.5.1 Spectroscopic characterisation of the monomer/dimer equilibrium

3.5.1.1 Dilution study of TF

The binding interfaces of proteins with small molecules or other proteins are regularly determined by NMR by means of fingerprint spectra, where each resonance corresponds to an isotopically labelled chemical group. Perturbations in the chemical shift across a titration of the substrate are mapped to particular residues involved in the interaction and in ideal cases can be used to determine a dissociation constant of the reaction. In a similar approach, ^{1}H-^{13}C HMQC spectra of TF^{ILV} were monitored across a range of TF concentrations. Dilution of the TF concentration perturbs the dimerisation equilibrium resulting in a gradual increase in the relative monomer population.

Experimental procedure : Two concentration titrations of TF were carried out : H_6-TF^{ILV} at 700 MHz and WT TF^{ILV} at 900 MHz. ^{1}H-^{13}C HMQC spectra were recorded of 7.5, 10, 15, 25, 40, 60 and 90 μM H_6-TF^{ILV} prepared by serial dilution, at 25°C and pH 7.5 at a magnetic field strength of 700 MHz. WT TF^{ILV} ^{1}H-^{13}C HMQC spectra were recorded at 4, 9, 13, 18, 26, 43, 65 and 87 uM at 25°C and pH 7.5 at a magnetic field strength of 900 MHz. The concentration dependent behaviour of representative resonances is shown in Fig 3.13. Resonance intensities across the two serial dilutions were compared after normalisation to the total TF concentration. Errors were calculated from the standard

deviation of the noise in individual HMQC spectra.

Predictions : Where chemical shift perturbations report on changes in chemical environment, resonance heights (as indirect measure of the linewidth) report on side-chain flexibility and the size of the overall molecule. Thus, in the case of TF dimerisation, where residues involved in the interface gain significant flexibility upon monomerisation in addition to an increase rotational correlation time, the resonance heights contain similar structural information as chemical shift perturbations. The dependence of the NMR signal on molecular size is discussed in more detail Section 2.1.2. In brief, the linewidth is equal to πR_2 and thus the height is proportional to $\sim 1/R_2$. The transverse relaxation rate R_2 is itself directly proportional to the rotational correlation time (τ_c) which is again related to the molecule size. Assuming a spherical particle, a doubling of the molecular weight results in doubling of τ_c and thus R_2. This increase in the R_2 results in the halving of the signal heights.

In this study, the monomerisation of the TF dimer results in two particles half the size of the dimer. Their associated resonance intensities, in the absence of exchange effects and

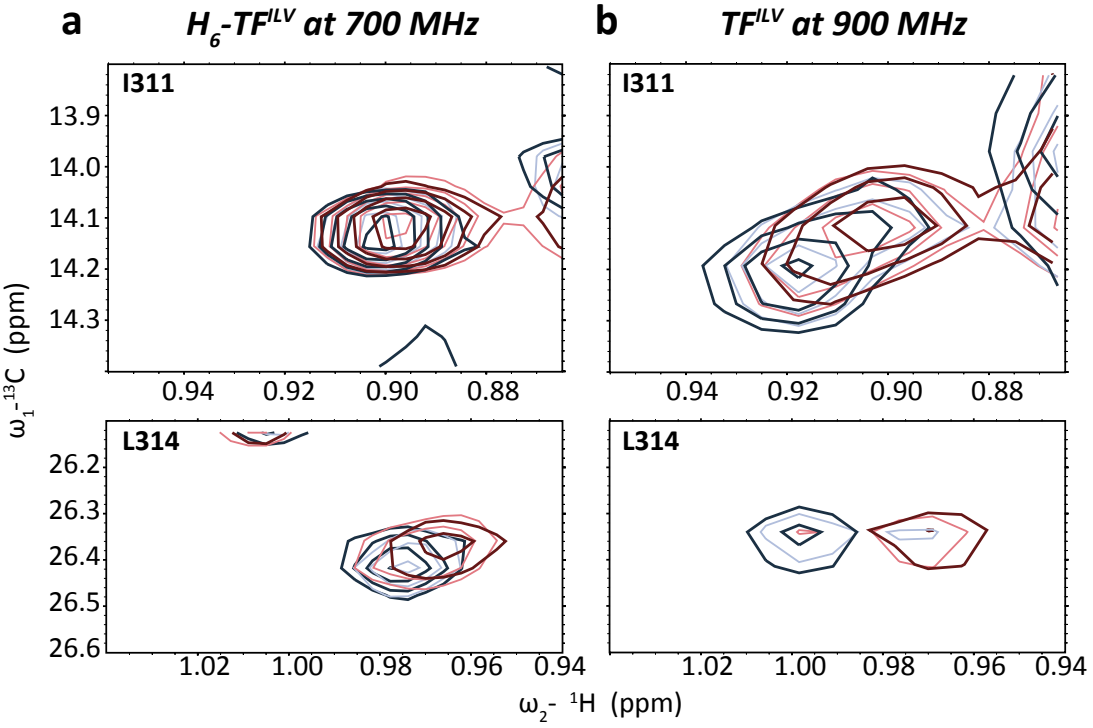

Figure 3.13: Exchange regimes of the TF dimerisation reaction.(**a**) Select H_6-TF^{ILV} ^{1}H-^{13}C HMQC resonances at 700 MHz : 7.5 μM in blue, 15 μM in light blue, 40 μM in pink and 90 μM in red. (**b**) Select TF^{ILV} ^{1}H-^{13}C HMQC at 900MHz : 4 μM in blue, 13 μM in light blue, 43 μM in pink and 87 μM in red.

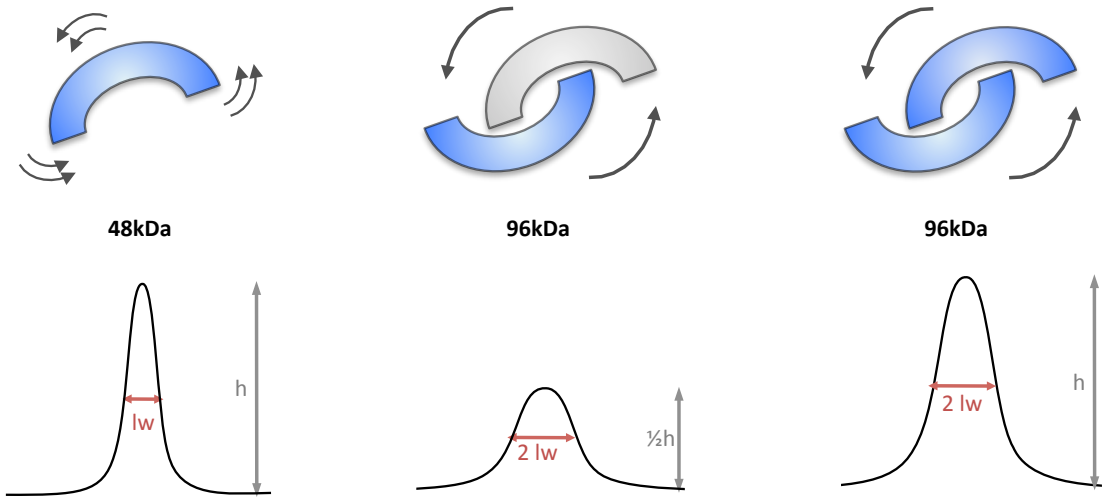

Figure 3.14: Size dependence of the detected NMR signal. The rotational correlation time ('tumbling') of small particles is rapid giving rise to narrow linewidths. For the same number of particles, a doubling in size gives rise to a doubling of the linewidth of the NMR signal. Since the integral of the signal is unchanged, the height of the associated peak is halved. Double the number of particles (represented here by a scenario where both monomers of the dimer NMR detectable (in blue)) results in doubling of the signal integral. However, since the size of the particle is conserved, the linewidth of the signal remains the same. The increase in the number of probes is therefor reflected in the signal height which is doubled, while the linewidth reflects the increase in molecule size.

thus not directly involved in the dimerisation process, should thus be double that of the same atoms in the TF dimer (Fig 3.14). Predictions about the TF population distribution over monomer and dimer states can be made using the mathematical formulation of the TF equilibrium described Section 3.4.2. Based on a literature dissociation constant value of 1.25 μM, 92% of TF will be dimeric at 90 μM and 68% and 75% only at 4 μM and 7.5 μM respectively. As the TF monomer contributes twice the intensity to the NMR spectrum, this change in the equilibrium is predicted to result in a ~1.2 fold increase of the detected resonance heights at 4 μM relative to the 87 μM spectrum of TF^{ILV} after normalisation for concentration. Based on a K_d value of 7.2 μM (see Section 3.5.1.2), the dimer population changes from 82% to 40% from 87μM to 4μM resulting in an expected 1.4 fold increase of the signal heights.

Dilution study of H_6-TF^{ILV} at 700 MHz : The dilution of H_6-TF^{ILV} resulted in progressive changes to the acquired ^{1}H-^{13}C HMQC spectra. Chemical shift perturbations were detected but were typically of small amplitude (~0.02 ppm in ^{1}H dimension, Fig 3.13.a). The majority of chemical shift perturbations mapped to residues in the TF PPIase domain. More extensive change to the resonance intensities were detected with decreasing TF concentration. The comparison of signal heights at 7.5 μM and 90 μM shows significant

increases in relative intensity upon dilution (Fig 3.15) up to 4 fold while only a limited number of resonances were broadened to 50%. The structural distribution of the observed relative intensities as presented in Fig 3.15.c displays a clear directionality. The limited number of broadened resonances map to the left-hand side of the protein in the orientation shown, covering all three domains. The more extensive resonance increase cover the right-hand side of the protein, in particular the arm 2 and neck regions of the SBD. In addition, while all 90 μM TF^{ILV} HMQC resonances are accounted for at lower concentrations (7.5 μM) novel resonances gradually appear with decreasing concentration.

The detected increase in the relative intensities as the monomer population progressively grows, with decreasing TF concentration, were above and beyond ($\sim$ 3 fold) what would be predicted based on the principles of size dependence of the NMR signal and predicted change in the TF population distribution as discussed above under slow exchange conditions (1.2-1.4 fold). The exchange between monomeric and dimeric TF thus appears to occur on an intermediate NMR timescale under the conditions of this study, with significant contribution from the exchange process to the transverse relaxation rates.

The study of the interaction of TF with α-synuclein described in Section 4.4.1 found the TF affinity was enhanced in the presence of a H_6-tag. The combination of the possibility of interference by the H_6-tag with dimerisation and the unfavourable exchange regime lead to the following study using the wild-type TF sequence and a higher magnetic field slowing down the relative timescale of the exchange (since the chemical shift difference increase with magnetic field strength)

Dilution study of WT TF^{ILV} at 900 MHz : The comparison of the ^{1}H-^{13}C TROSY HMQC at 4 μM and 87 μM is shown in Fig 3.16.a. Significant, yet small, differences in chemical shifts (10-15 Hz) and more substantial changes in resonance intensities were detected. Chemical shift perturbations map predominantly to the PPIase domain and the arm 1 region of the SBD.

The peak heights at 4 μM relative to 87 μM were normalised for concentration and shown in Fig 3.16.c. Selective broadening of resonances to 40% was observed, mapping in particular to the RBD, arm 2 and cavity regions of the SBD of TF. A limited number of resonances originating from the neck region of the SBD displayed significantly increased intensities between 150-200%. While the expected increase in signal intensity discussed above (1.2-1.4 fold) is observed for a minority of probes, in particular between 150 and 250 (i.e the PPIase domain), a large number of resonances are significantly reduced to about 50%. The observed reduction reflects the slow exchange behaviour where the conversion from the dimer to the monomer species results in a decrease of the dimer resonance monitored here and an increase in the monomer signal elsewhere. Since this exchange phenomenon can only be observed if the chemical shift between the two species differs,

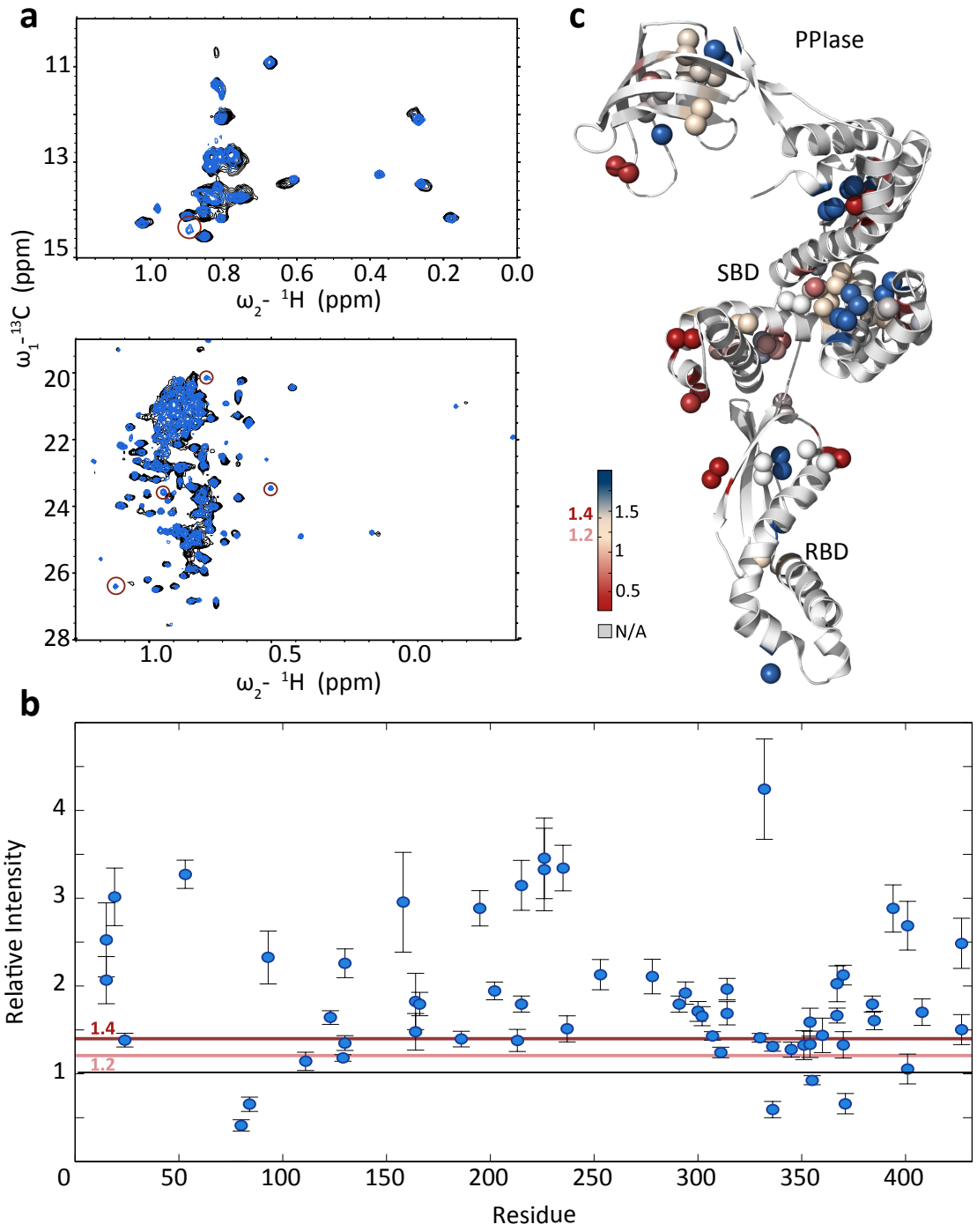

Figure 3.15: Identification of residues involved in dimerisation through ^{1}H-^{13}C TROSY HMQC resonance heigths. (**a**) Ile and Leu/Val region of the ^{1}H-^{13}C TROSY HMQC spectra of 7.5 μM (black) and 90 μM (blue) H_6-TF^{ILV}. (**b**) Sequence distribution of ^{1}H-^{13}C TROSY HMQC resonance intensities at 7.5 μM relative to 90 μM. Pink and red lines represent expected intensity values in the absence of exchange effects for a dimerisation K_d of 1.25 and 6.8 μM respectively. (**c**) Structural distribution of observed intensity changes as shown in b.

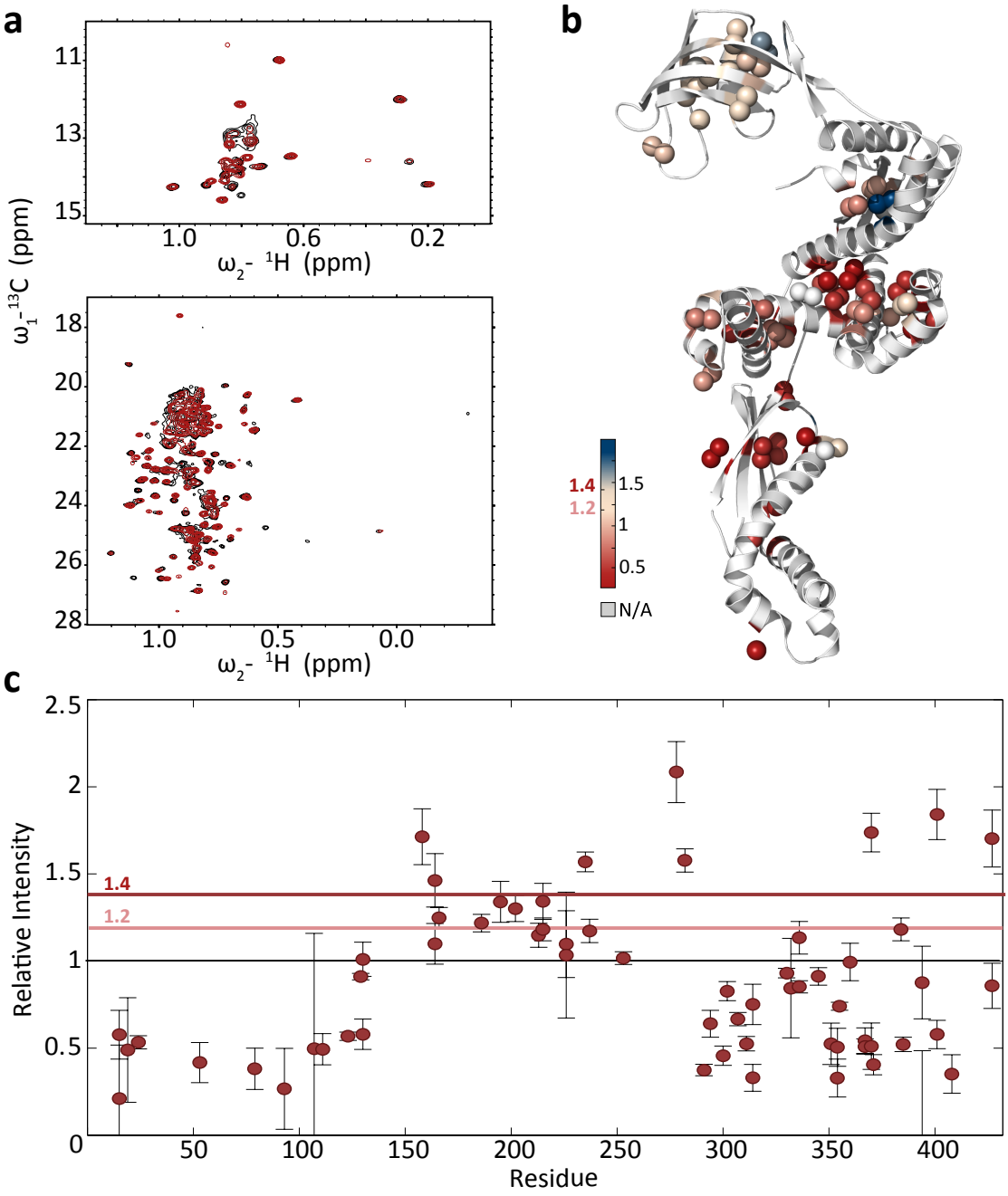

Figure 3.16: Identification of residues involved in dimerisation through ^{1}H-^{13}C TROSY HMQC resonance heigths. (**a**) Ile and Leu/Val region of the ^{1}H-^{13}C TROSY HMQC spectra of 4 μM (black) and 87 μM (red) TF^{ILV}. (**b**) Structural distribution of observed intensity changes as shown in c. (**c**) Sequence distribution of ^{1}H-^{13}C TROSY HMQC resonance intensities at 4 μM relative to 87 μM. Pink and red lines represent expected intensity values in the absence of exchange effects for a dimerisation K_d of 1.25 and 6.8 μM respectively.

this indicates that the chemical environment of the chemical group in question in the dimer and monomer is significantly different.

The intensity data here presented thus implicates both the full RBD and SBD in the dimerisation reaction based on the selective reduction in resonance heights. From this analysis, the PPIase domain appears unaffected by monomerisation. The present analysis does however not account for resonances in the fast exchange regime. Taking into consideration that the majority of resonances in the TF^{ILV} HMQC spectra presented here are in the slow exchange regime, with relative small chemical shift differences (10-15 HZ), the chemical shift difference between monomer and dimer states would have to be < 1Hz to put the resonance in the fast exchange regime (Section 2.1.6.1), assuming the exchange rate of dimerisation is a global parameter.

Discussion Both the titration of H_6-TF^{ILV} at 700 MHz and WT TF^{ILV} at 900 MHz identified the RBD and arm 2 region of the SBD as most strongly affected by changes in the monomer/dimer equilibrium. A majority of TF resonances at 700 MHz were found to be in intermediate exchange. The increase in field strength to 900 MHz effectively slows down the relative exchange timescale and most resonances were in slow exchange in the second dilution series described. A principle difference between the two data sets is the role of the PPIase. The titration of H_6-TF^{ILV} found a number of residues involved in the dimerisation process, while at 900 MHz, WT TF resonances from the PPIase domain behave according to the predictions made at the start of this study. While the exact origin of the observed differences remains unclear, the H_6-tag may perturb the dimerisation process slightly.

It should further be noted that at high concentrations of WT TF, only one set of resonances was observed. The total number of detected resonances as described Section 3.3.4 (126 out of total of 171 I, L and V methyl groups present in TF) indicates that both monomers contribute to the same dimeric signal. It is therefor likely that the TF dimer interface is symmetrical.

3.5.1.2 Numerical analysis of the WT TF dilution data set

The TF^{ILV} dilution data set displays favourable slow exchange behaviour (Fig 3.13) for the quantification of the concentration dependent changes observed. In this slow exchange limit, the shift in the population distribution upon dilution thus results in the gradual disappearance of the dimer resonances and a gradual increase in the monomer contribution to the spectrum.

K_d determination by two-dimensional line-shape analysis : A value of the dissociation constant for the monomer/dimer equilibrium was obtained from the WT TF^{ILV} titration series by means of a two dimensional line-shape analysis approach (Waudby et

al, in preparation). Lineshape analysis monitors the chemical shift and linewidth of an NMR resonance across a titration in order to determine the K_{d} and kinetic details of the reaction as described by the BlochMcConnell equations and is therefore not limited to any particular exchange regime. Preliminary analysis of two resonances pairs (fits of I311 presented in Fig 3.17.d) produced a K_{d} value of 7.2 μM and a dissociation rate significantly slower than 1 s^{-1}. The inclusion of more pairs of resonances should increase the robustness of the obtained results but also forms the main caveat of this approach as partner resonances (originating from the same chemical group in the monomer and dimer) need to be identified with certainty.

Global fit of the dissociation constant of dimerisation by linear regression : Using a linear regression approach developed in collaboration with Dr Chris Waudby (UCL), signal intensities at each point of the full HMQC spectra were decomposed into monomer and dimer contributions identified by concentration dependent changes in the signal intensity. The full Matlab script is included in Appendix A.4-5. Expected concentration dependent behaviour was predicted for a range of dissociation constant values and optimal values of I_{dimer} and I_{monomer} determined by fitting of the experimental data. The deconvoluted monomer and dimer signal were recombined according to the predicted population distribution and compared to the experimental data to calculate residuals and hence Chi^2 values for every K_{d} value screened.

Fit results are shown in Fig 3.17.a. Preliminary results demonstrated that certain sites were fitted with negative signal contributions from one of the two species present and signals were therefore restrained to be positive. Without this physically-motivated constraint, lower K_{d} values were found to be preferred. Initial fits were performed across all points of the TF HMQC spectrum. However, most of these points contain noise only and thus do not contain information about the dissociation constant of the dimerisation reaction, potentially introducing errors in the obtained fitting results. Therefore, a signal-to-noise threshold was applied, excluding the majority of the noise. The effect of the S/N threshold on the fitting results and the portions of the NMR spectrum included in the fit are shown in Fig 3.17.a and c. More stringent filtering of the signals included in the fit resulted in marginally lower K_d values (from 11.6 μM without threshold to 10 μM with a threshold of 10 S/N) but all values were within error, proving that the linear regression approach detailed here is robust. A signal-to-noise threshold of 3 was chosen for the final determination of the dissociation constant as it efficiently filtered out all noise while including weaker TF^{ILV} signals (Fig 3.17.c). The final dissociation constant obtained by this method is 10.5 $\pm$ 0.4 μM. Errors were estimated from the Chi^2 analysis, where in the case of one single degree of freedom in the fit (here, the dissociation constant) a value of Chi^2=1 corresponds to the upper and lower limits defined by the experimental error.

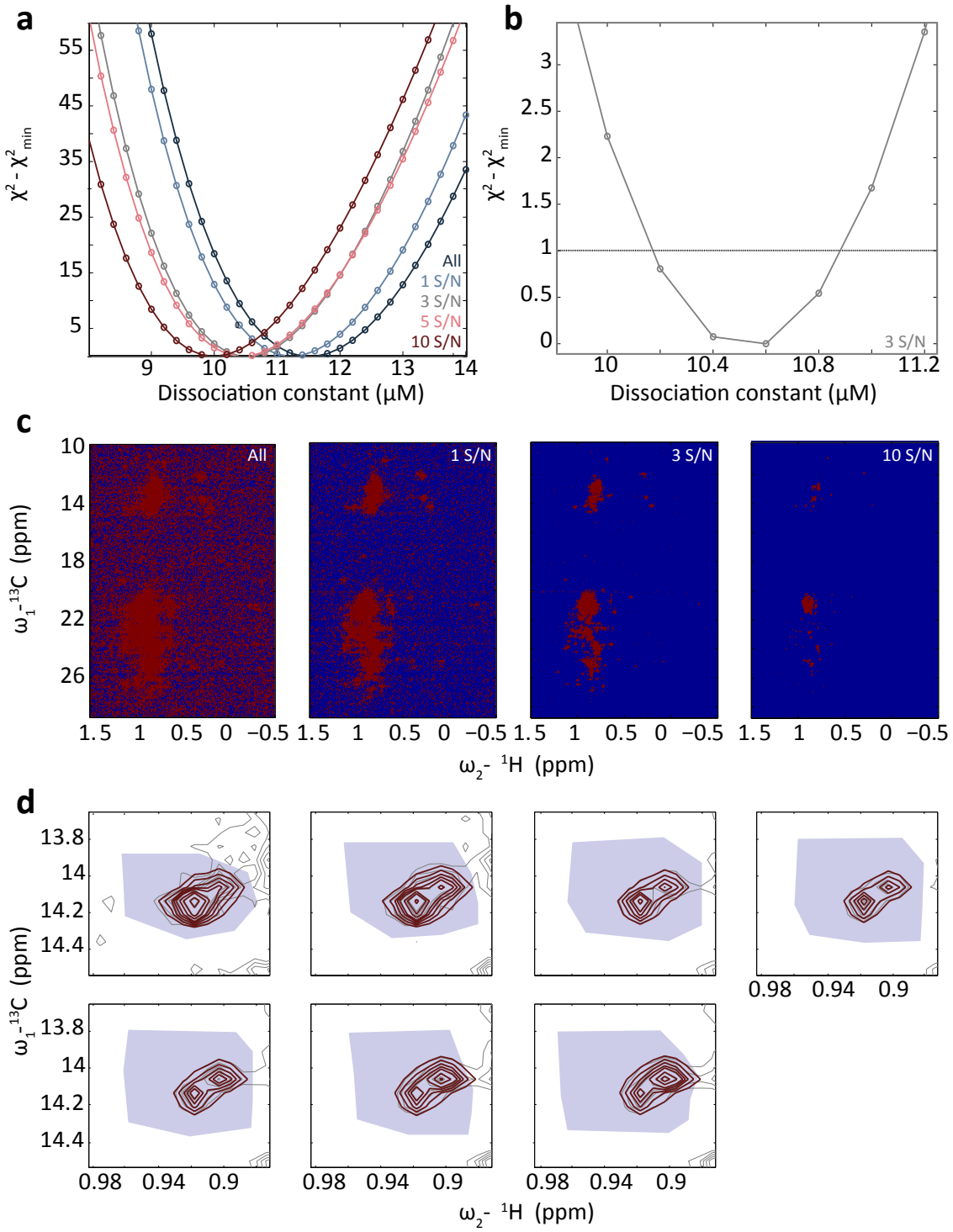

Figure 3.17: Determination of the dissociation constant of TF dimerisation. (**a**) Effect of data threshold on fitted dissociation constant. (**b**) Chi^2 minimisation of the TF dimerisation dissociation constant with a inclusion threshold of 3 S/N. Enlargement around minimal Chi^2 value of curve shown in a. (**c**) Effect of threshold on the physical appearance of the ^{1}H-^{13}C TROSY HMQC spectrum. (**d**) Overlay of simulated NMR data based on lineshape analysis (red) and the experimental spectrum (grey) for residue I311 across the 7 concentrations of the titration Section 3.5.1.

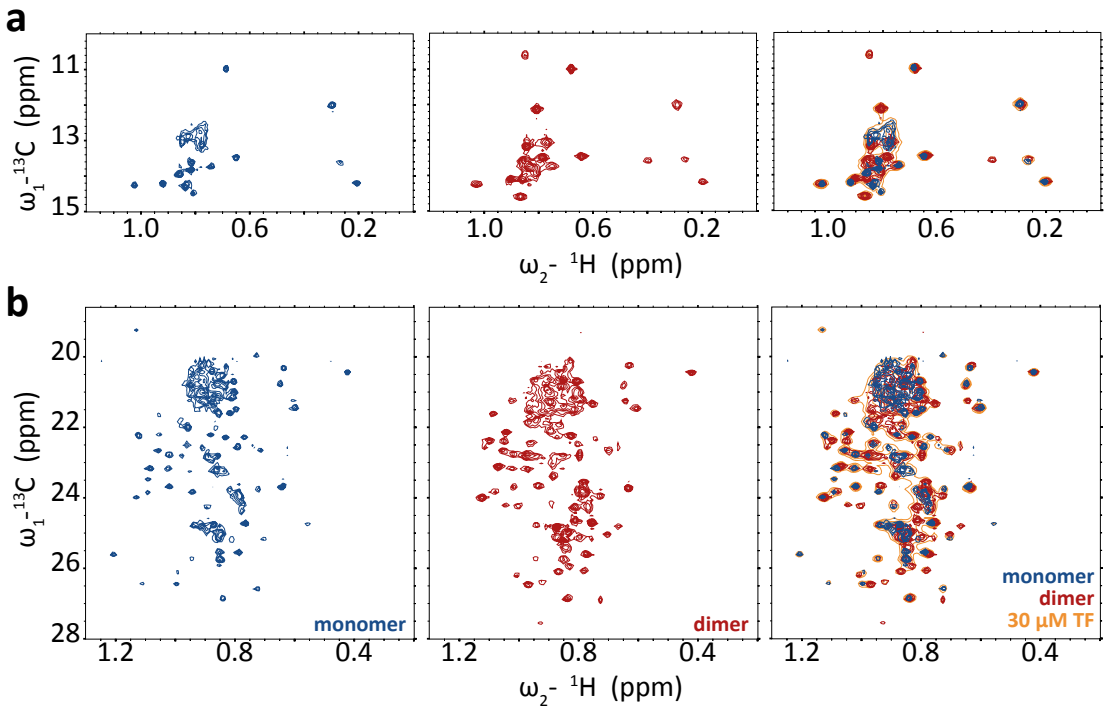

Figure 3.18: Simulation of NMR data based on deconvoluted monomer and dimer contributions to the TF ^{1}H-^{13}C TROSY HMQC Ile region (**a**) and Leu/Val region (**b**) of deconvoluted TF monomer (blue) and dimer (red) NMR spectrum. The 30 μM TF^{ILV} spectrum (yellow) is the linear combination of the monomer and dimer spectra according to the TF population distribution.

3.5.1.3 Deconvolution of the monomer and dimer contributions to the TF^{ILV} NMR spectrum

The numerical analysis described in Section 3.5.1.2 decomposed observed spectra into monomer and dimer contributions. Based on the best-fit K_{d} of 10.5 μM for dimerisation, the NMR spectra of 1μM and 1μM pure monomer and dimer were simulated as shown in Fig 3.18. Overlay of the experimental TF^{ILV} spectrum at 30 μM with the monomer and dimer spectra shows that the experimental data is indeed a linear combination of the two simulated spectra. Since the monomer and dimer signals evolve in the opposite direction with increasing TF concentration, this deconvolution of the monomer and dimer contributions proved highly robust to the assumed K_{d} value for the dimerisation reaction.

This linear regression method has thus permitted the determination of the NMR spectrum of a pure species that is in theory inaccessible to NMR: the TF monomer. This approach is particularly useful for those resonances with small chemical shift difference across the two species. While in the experimental data, the two resonances overlap and thus appear as a single, broad, resonance, this deconvolution highlights the two different

contributions and demonstrates that even for resonances that visually do not appear to change significantly across the TF concentration titration such as V215 and V164 (TF assignment in Section 3.3.5), the NMR signals originating from the monomer and the dimer may in fact be distinct.

The deconvolution of the monomer and dimer contributions further permits the prediction of the appearance of the TF NMR spectrum given a certain population distribution as demonstrated for the TF spectrum in the presence of the ribosome in Section 4.5.

3.5.2 Structural characterisation of the TF dimer

The analysis of TF^{ILV} resonance intensities across a serial dilution (Section 3.5.1.1) implicated a large region of the RBD and arm 2 and cavity regions of the SBD in the dimerisation reaction. NMR signal heights are sensitive reporters of changes in the chemical environment and thus reflect both residues in the dimer interface and associated conformational changes in more distal regions of the protein. In order to narrow down the exact interface residues, cross-saturation experiments were pursued. Paramagnetic relaxation enhancement measurements were undertaken to provide additional medium range distance restraints in order to determine the relative orientation of the two TF molecules in the dimer.

3.5.2.1 Cross-saturation mapping of dimer interface

The observed changes in peak height and chemical shifts across concentrations as described in Section 3.5.1 reflect any changes in the chemical environment or flexibility of I, L and V side-chains upon monomerisation. Perturbed resonances thus do not uniquely map to the dimerisation interface but may also reflect more widespread conformational changes associated with dimerisation. Cross-saturation experiments (Section 2.1.6.4) were therefore used to specifically identify interface residues.

80 μM [^{2}H,^{13}C-ILV]-TF was mixed with 320 μM [^{1}H,^{12}C]-TF such that approximately 80% of the total TF^{ILV} present forms mixed ^{1}H/ILV dimers. TF^{ILV} ^{1}H-^{13}C HMQC spectra were then recorded at 25°C, pH 7.5 at 700MHz with an recovery delay of 1.2 s. Spectra were recorded with and without selective ^{1}H saturation pulse at 4.13 ppm during the recovery period and the relative peak heights were then compared. Overlay of the TF^{ILV} ^{1}H-^{13}C TROSY HMQC spectra recorded shows certain resonances were selectively broadened (Fig 3.19.a). Uniform broadening to 90% was observed along the sequence indicating either that the selective saturation of protons was not entirely limited to the protonated TF monomer (Fig 3.19.b) or spin diffusion across the whole molecule. Additional, selective, broadening at sites such as between residues 50 and 100 indicate saturation transfer across the dimer interface.

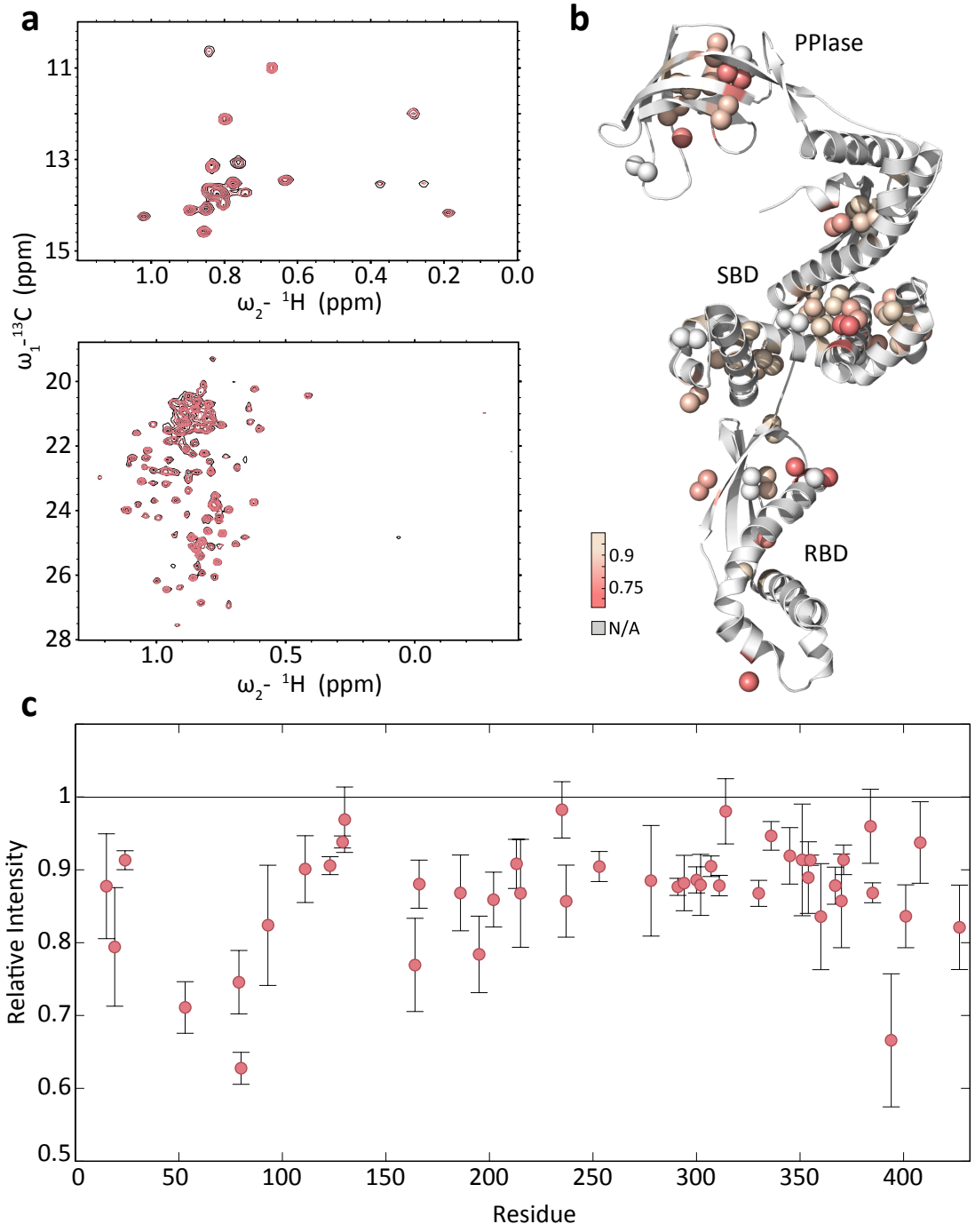

Figure 3.19: Dimerisation interface residues based on cross-saturation NMR experiments. (**a**) Ile and Leu/Val region of the ^{1}H-^{13}C TROSY HMQC spectra of TF^{ILV} in the absence (black) and presence (pink) of selective ^{1}H saturation pulse. (**b**) Structural distribution of observed intensity changes as shown in c. (**c**) Sequence distribution of ^{1}H-^{13}C TROSY HMQC resonance intensities in the presence of selective ^{1}H saturation pulse relative to intensities in the absence of saturation pulse.

The observed broadening maps to residues in the extremity of 'arm 2', the inner surface of the RBD domain and some residues in the PPIase domain (Fig 3.19.b). The majority of assigned I, L and V resonances in the RBD display relatively strong broadening by saturation transfer, although given the limited number of probes in this domain, this does not provide a clear picture of the exact outline of the interaction surface. The main identified sites overlap with those most strongly affected by dilution (Section 3.5.1.1) with the exception of the PPIase domain where little to no change was observed by monitoring the resonance intensities (Fig 3.16).

The cross-saturation data presented here show that residues on the inner surface of TF (facing upwards in Fig 3.19.b) line the dimerisation surface. These include residues identified as involved in substrate binding [130]. The absence of broadening detected further into the cradle suggests that packing of the dimerisation surface is not very tight. The dimerisation interface further appears to be a discontinuous surface mapping to separate sites on individual domains.

3.5.2.2 Structural characterisation of the TF dimer using intermolecular paramagnetic relaxation enhancements

Cross-saturation and intensity analysis have shown that interface residues localise to the main hydrophobic pockets present on the TF surface (Fig 1.19) alluding to the nature of the contacts between monomers, but these measurements do not provide information on the relative orientation of the sub-units in the dimer. Medium range distance constraints (10-25 Å) from spin-labelled positions to assigned ile, leu and val methyl groups in the TF dimer were therefore measured by paramagnetic relaxation enhancement experiments (Section 2.1.6.5) in order to obtain a more detailed description of the arrangements of the TF domains relative to each-other within the dimer.

Experimental strategy and procedure : Proxyl spin-labels were introduced at 4 distinct sites using single cysteine mutant, [^{2}H]-TF at positions 14, 150, 326 and 376 (Section 2.2.5.3). 150 μM spin-labelled TF mutant was mixed with 50 μM TF^{ILV} and recorded at 25°C, pH 7.5 at 800 MHz. ^{1}H-^{13}C HMQC spectra of both a paramagnetic and diamagnetic sample were acquired with additional relaxation delays of 0.1 ms and 20 ms respectively. Diamagnetic samples were prepared by addition of 2 mM ascorbic acid. The reduction reaction was allowed to proceed for 1 hour before further NMR data was acquired. NMR spectra were normalised for dilution by comparison of the DSS signals in the paramagnetic and diamagnetic samples.

While ^{1}H-^{13}C HMQC spectra were acquired at two relaxation delays as required for the precise determination of distances from PRE data using the analysis described in Section 2.1.6.5, the weak PRE observed and the limiting effect of the TF dimer size on the

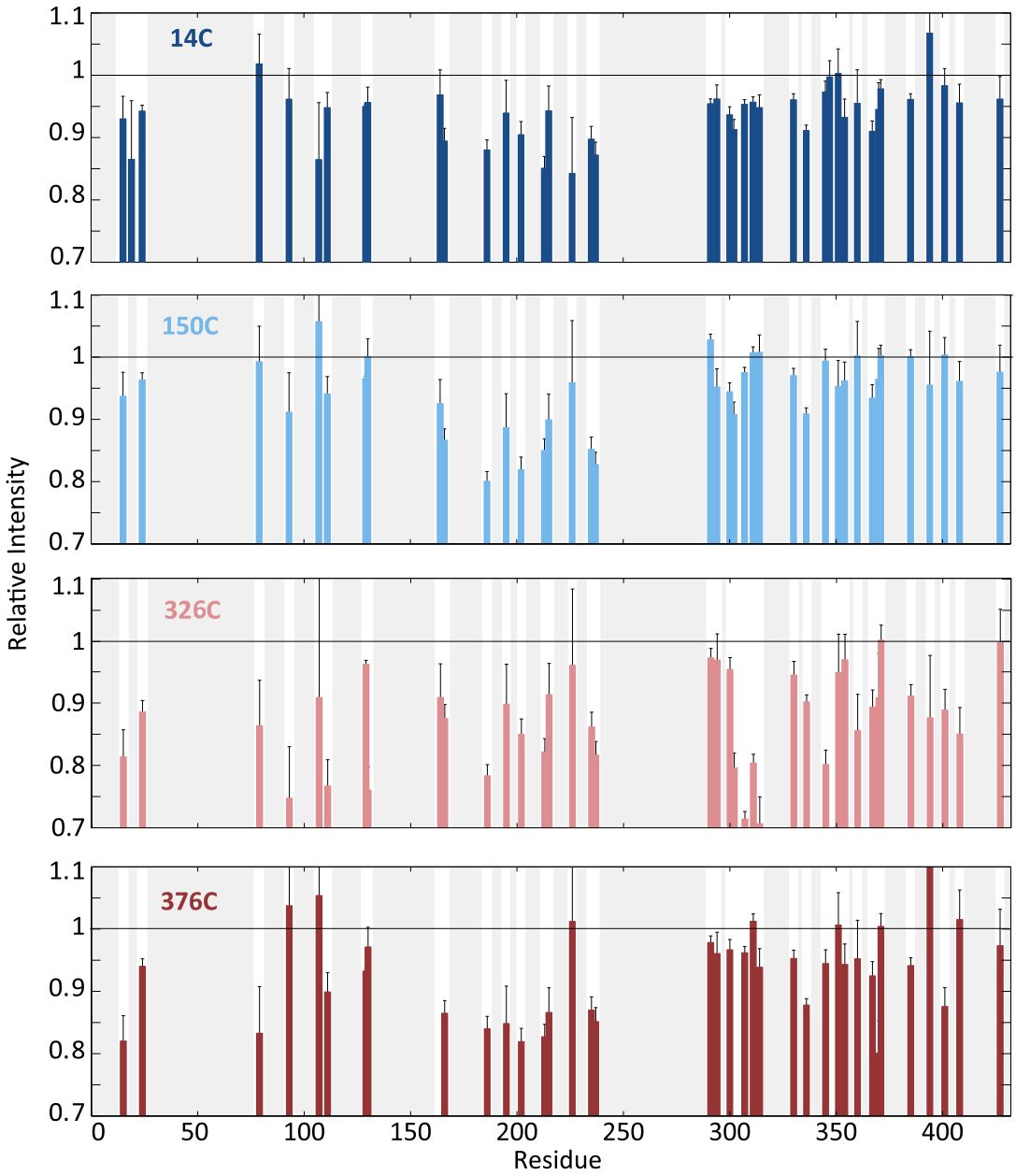

Figure 3.20: Effect of PRE on TF ^{1}H-^{13}C HMQC resonance intensities. Paramagnetic ^{1}H-^{13}C HMQC TF^{ILV} intensities relative to diamagnetic intensities for 50 μM TF^{ILV} in the presence of 150 μM [^{2}H]-TF spin-labelled at positions 14, 150, 326 and 376 at 25°C and 800 MHz.

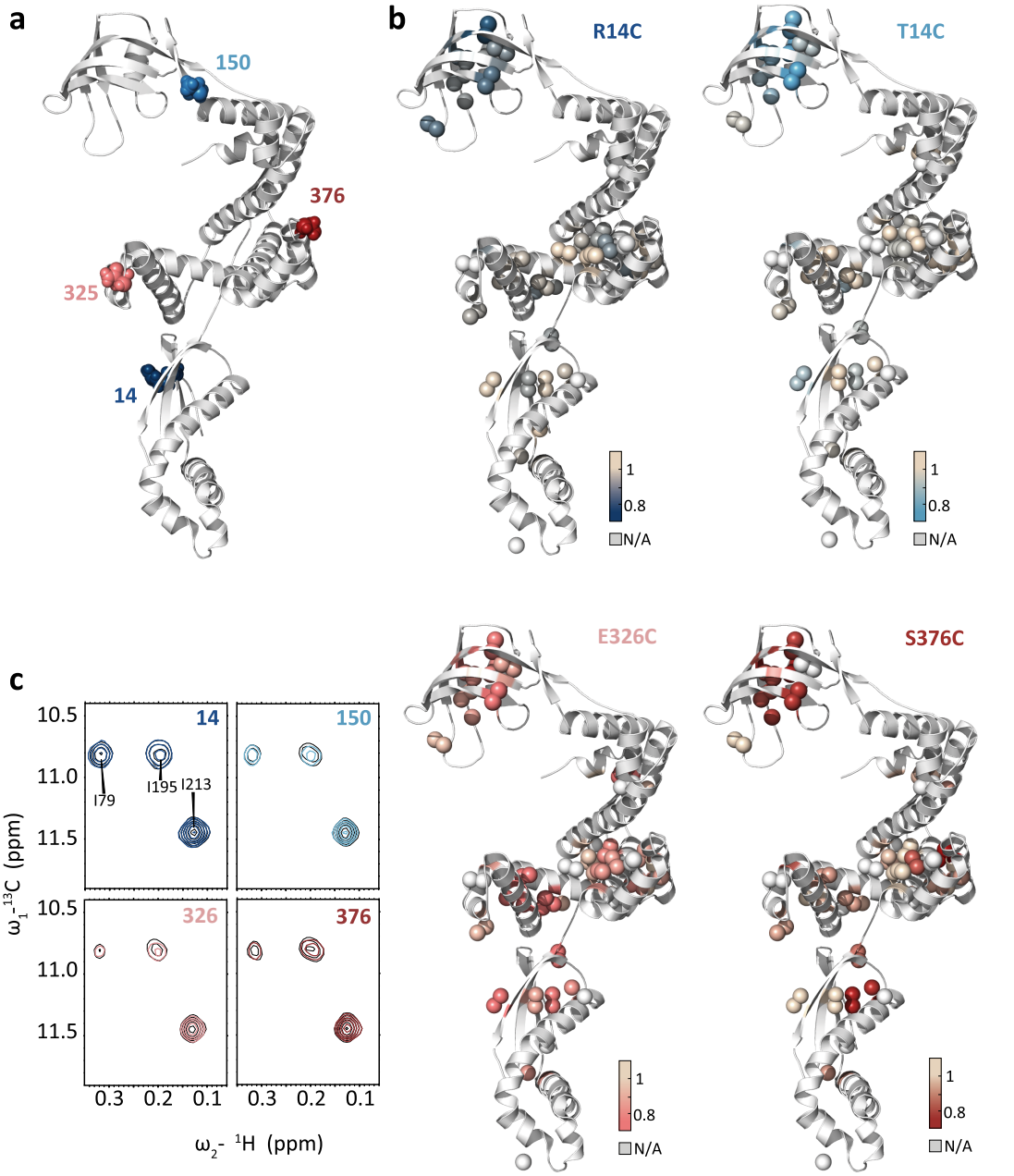

Figure 3.21: Structural distribution of measured paramagnetic relaxation enhancements. (**a**) Sites of spin-label attachment: 14 (darkblue), 150 (blue), 326 (pink) and 376 (red). (**b**) Structural distribution of experimental PREs shown in Fig 3.20. Colouring as in a. (**c**) Region of the ^{1}H-^{13}C HMQC spectra of paramagnetic and diamagnetic TF samples.

quality of the NMR data of TF did not permit the accurate determination of relaxation rates. Instead, resonance heights in the short delay (0.1 ms) ^{1}H-^{13}C HMQC spectra (where signal is strongest due to the shorter relaxation time), were compared (Fig 3.20) across the diamagnetic and paramagnetic samples as a semiquantitative indication of proximity to the spin label.

Observations and discussion : Weak PREs (I/I_0 ∼0.85) were observed for spin label positions 14 and 150 (Fig 3.20). In both cases, broadening appears limited to the PPIase domain (Fig 3.21) with, minor additional perturbations observed in the arm 1 region of the SBD for position 14. Significantly more signals were perturbed by the presence of spin-labels at positions 326 and 376. The proxyl spin-label at position 326 caused the largest perturbation observed down to 70% the intensity in the diamagnetic sample. All TF domains were affected with an apparent ranking from strongest to weakest : the RBD and arm 1 region (I/I_0 ∼0.75), the PPIase domain and the arm 2 region (I/I_0 ∼0.85). PREs from position 376 were mainly found in the RBD and PPIase domain (I/I_0 ∼0.85) with weaker effects observed in the arm 2 region of the SBD ($I/I_0 \sim 0.9$).

The limited PRE effect detected indicates all spin-labelled positions are relatively far removed from the dimer interface. Indeed, nitroxide spin PREs have been detected up to 25 Å while distances of 5 Å typically result in complete broadening of NMR signals. From the TF crystal structure [126] (Fig 3.21.a), all spin-label sites appear to point away from the sites identified in Sections 3.5.1.1 and 3.5.2.1 to be involved in dimerisation. The broadening observed for all spin-label positions in resonances mapping to the PPIase domain is remarkable since the spin-label positions cover the full monomer structure. It suggests that the PPIase domain in the solution TF dimer, undergoes a significant conformational change relative to the dimer crystal structures (Fig 3.12).

3.5.2.3 Model of the solution structure of the TF dimer

An initial model of the solution structure of the TF dimer can be proposed based on the dilution, cross-saturation and PRE experiments described in this Section. The model is presented in Fig 3.22 and displays a symmetrical head-to-toe conformation. Main dimerisation contacts involve the RBD and the extremities of the SBD arm regions. Minor contacts are formed by the PPIase which is bend over the second sub-unit's RBD both occluding the ribosome binding motif and adopting a more central position to satisfy the observed PRE distance restraints. Within this conformation, the identified substrate interaction sites are occluded and partially involved in the dimerisation interface itself. The widespread intensity changes in the dilution study indicate that dimerisation is likely paired with conformational changes in the relative orientation of the domains in particular the RBD and SBD domains. While the location of the PPIase domain does not reflect a

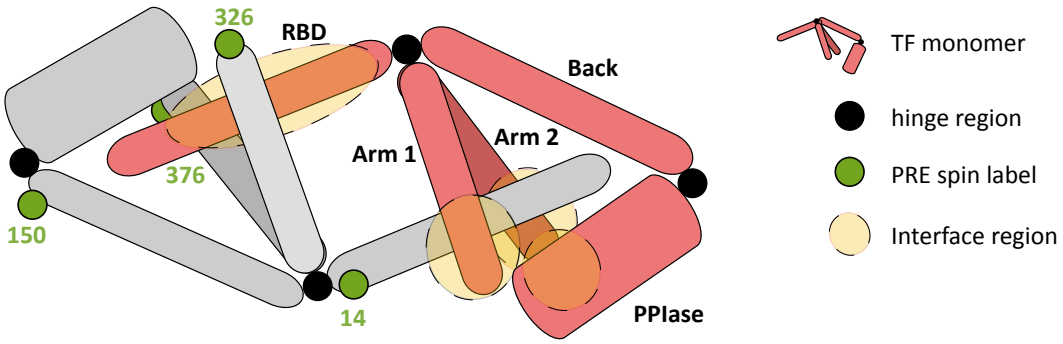

Figure 3.22: Model of the solution structure of the TF dimer

previously observed conformation, significant flexibility of the neck hinge region has been proposed based on the various orientations adopted across different TF monomer crystal structures as described in further detail in Section 3.5.3.

The structural data presented here is currently being used as conformational restraints in computational simulations using the HADDOCK webserver [244, 245] in order to build a high resolution model of the TF dimer structure in solution. Significant conformational changes relative to the available monomer structures appear required in order to satisfy all available constraints. As a consequence, efforts thus far have not yet resulted in a satisfactory model. Alternative computational approaches that may be pursued to account for the apparent dynamics within the TF dimer are restrained molecular dynamics (MD) based on simulated annealing protocols or *ab initio* NMR structure calculation software such as ARIA. Since full atom simulations of systems as large as the TF dimer are computationally expensive, ideally, an initial restrained docking step based on all described NMR data would lead into a structure optimisation, and visualisation of dynamics by MD.

3.5.3 Characterisation of dynamics in the trigger factor dimer

The TF monomer structure has been extensively studied at high resolution yielding a number of crystal structures from various species and in various macromolecular assemblies (Fig 3.23). Although the topology of TF is conserved across all published structures, the relative arrangements of domains vary with significant rotational freedom around two key hinge regions connecting the RBD and SBD and SBD and PPIase domains respectively. In particular the PPIase domain appears to possess considerable freedom to reorientate relative to the SBD domain. Further differences observed between the monomer structures involves the widening of the distance between the two arm domains in particular in the presence of a substrate protein. Finally, the RDB was shown to undergo a significant conformational change upon interaction with the ribosome surface resulting in the exposure

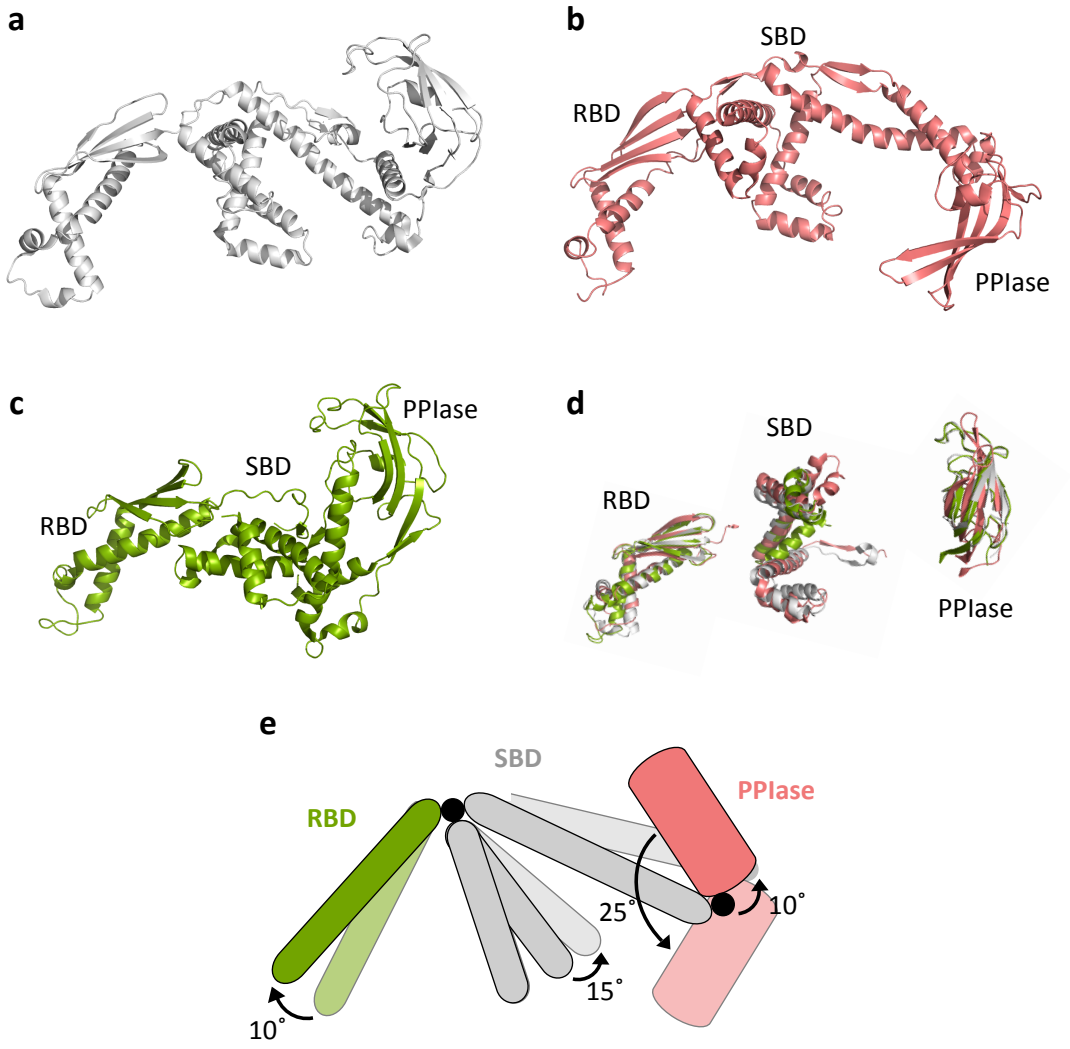

Figure 3.23: Structural heterogeneity in the TF monomer (**a**) Crystal structure of *Ecoli* TF (pdb 1W26 [126]). (**b**) Crystal structure of *T.maritima* TF (pdb 3GTY [84]) (**c**) Crystal structure of *V.cholerae* TF (pdb 1T11 [238]). (**d**) Overlay of individual domains of the three afore structures shows structures are highly conserved. (**e**) Model of the differences in domain orientations across the three crystal structures and S7-bound structure.

of a more hydrophobic surface [135].

A detailed characterisation of the dynamics of the TF molecule could provide insights into the relation between the numerous crystal structures and explain their physiological relevance in the TF chaperoning mechanism. NMR spectroscopy can provide detailed dynamic information on many physiologically important timescales. Here, pico to nano second dynamics were evaluated by means of order parameter measurements as indicators

of the structural stability of TF. Secondly, we have investigated chemical exchange processes occurring on a millisecond timescale to understand slower motions within the TF molecule that may be associated with dimerisation.

3.5.3.1 Picosecond-nanosecond timescale dynamics in the TF dimer

An NMR strategy outlined in [211], allows the measurement of order parameters describing the amplitude of motion of the methyl 3-fold symmetry axis in selective methyl protonated samples. Fast motions in the methyl symmetry axis report on the extent of burial of a methyl group within the protein structure. Since methyl-containing residues are mainly hydrophobic, they can typically be found buried in folded cores of proteins. Rapid motions in the methyl axis are therefore indicative of 'breathing' of the protein core or rapid unfolding and refolding of certain structural elements. Care should be taken in the attribution of low order parameters to such events as exposed (non buried) methyl groups will display similar behaviour without reporting on local dynamics, in particular in long side-chain residues. The order parameters of methyl groups are therefore both structural and dynamic reporters. Where the tertiary structure is known, as is the case for TF, the two contributions can be distinguished to some extent.

Methyl axis order parameters are determined from the ratio of double quantum coherence build-up and bi-exponential relaxation of transverse ^{1}H magnetisation intensities recorded for various relaxation delays as described in Section 2.1.6.3. Experimental details are further described in Section 2.1.6.3. Values of the order parameter for each resonance are calculated from equation 2.1.9, Section 2.1.6.3. Equation 2.1.10 is fitted to the 12 point experimental data in Matlab and describes the relationship between the intensity ratio of the two experiments, the intra-methyl ^{1}H-^{1}H dipole cross-correlated relaxation rate (η), a measure of the proximity of external (not belonging to the studied methyl group) protons (δ) and the relaxation delay (T). The intra-methyl ^{1}H-^{1}H dipole cross-correlated relaxation rate can then used to determine the order parameter values from equation 2.1.9. While the rotational correlation time of the TF dimer has not been experimentally derived, a value of 60 ns was chosen based on the approximation that the correlation time in ns corresponds to approximately 1/2 the MW in kDa for a spherical particle. This estimate is validated by the observation that all experimentally derived values of S^2 were between 0 and 1.

Order parameter values were determined for 34 out of 51 assigned TF residues (Fig 3.24). The determination of accurate S^2 values was hindered by overlap in the methyl TROSY HMQC spectrum of TF^{ILV} preventing accurate determination of peak intensities and low signal to noise in the double-quantum build-up experiments. Residues for which the order parameters could be accurately measured display overall high values (>0.6),

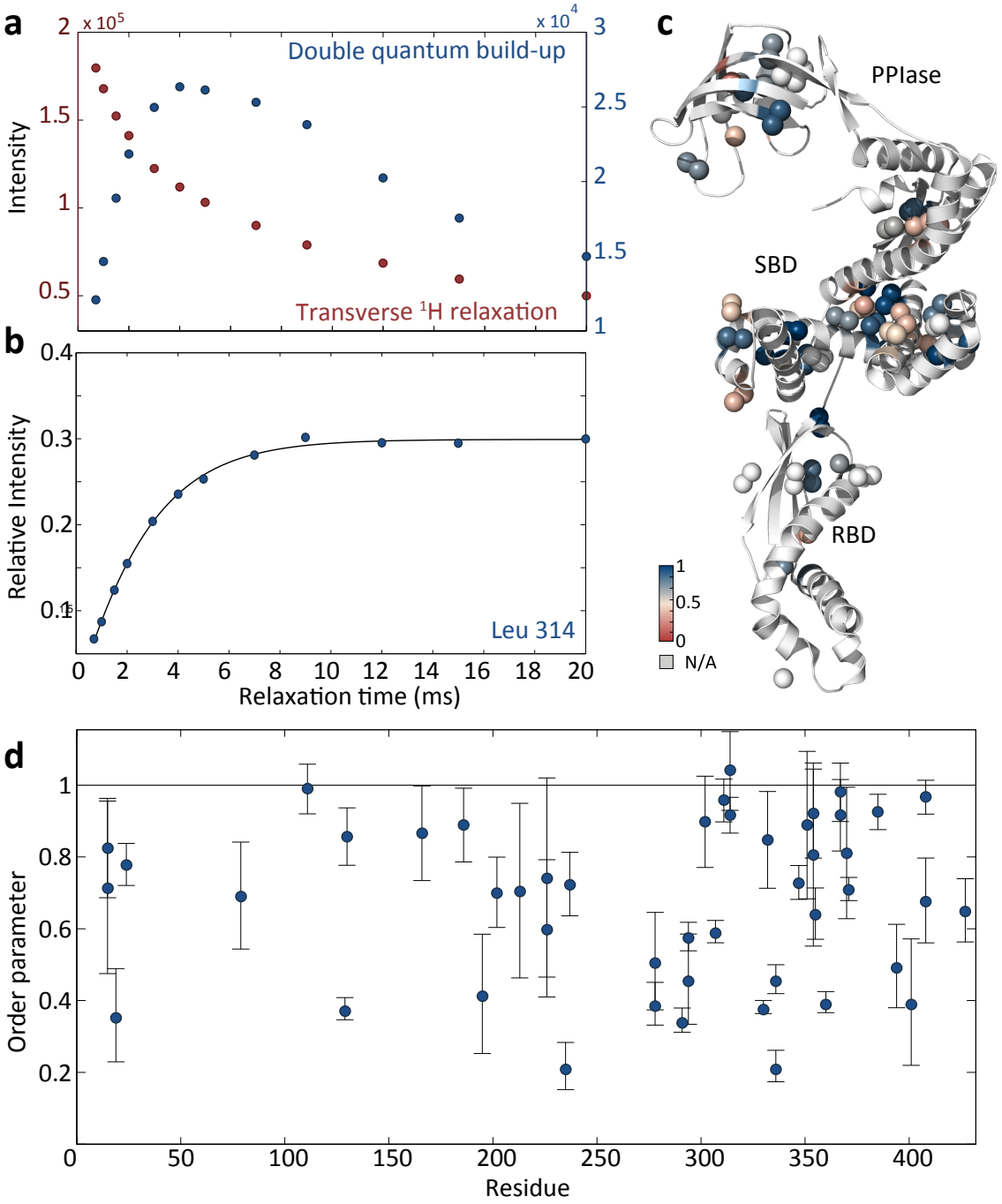

Figure 3.24: Pico to nano second dynamics in TF. (**a**) Absolute L314 resonance intensities in function of relaxation times in the double-quantum build-up and transverse relaxation experiments. (**b**) Fit of the ratio of the intensities in as described in text. (**c**) Structural distribution of determined order parameters from 0 (red) to 1 (blue). (**d**) Sequence distribution of experimentally determined order parameter values of side methyl groups of ile, leu and val residues in TF.

indicating high order. The main structural features of TF are thus overall rigid on a pico-nano second timescale and most TF methyl groups are buried within structural elements confirming that the selected residue types: ile, leu and val provide valuable probes of structural changes within TF. The observed high order parameters are in agreement with expected values for a well-folded protein and demonstrates the structural integrity of all the TF structural elements. A small number of residues were found to exhibit faster pico to nano second scale dynamics (Fig 3.24) without clear indication from the crystal structure that these residues are exposed on the protein surface. They appear to map to the interfaces of secondary structural elements indicating possible fluctuations in the relative arrangement of structural entities within domains without disrupting the main protein fold.

3.5.3.2 Millisecond timescale dynamics in the TF dimer

Experimental details : ^{1}H-^{13}C-MQ-CPMG relaxation dispersion experiments (Section 2.1.6.2, [182]) were recorded of concentrations between 200-800 μM TF^{ILV} at pH 7.5 in 100% deuterated Tico buffer at 15, 20 and 25 °C at 700 and 900 MHz magnetic field strength and at 25 °C only at 500 and 800MHz. The total CPMG delay was optimised to 20 ms for an optimal balance of signal and minimum CPMG frequency, and 16 CPMG frequencies were used between 0 and 1000 Hz. Further experimental details are provided in Section 2.3.2.8. 64 resonances were fitted in Fuda across the CPMG frequencies and analysed using Matlab based program GUARDD [234].

Exchange contribution to TF relaxation rates. Fitted resonance intensities were converted to apparent relaxation rates and visually inspected for dispersions across CPMG frequencies. 44 resonances were found to display dispersions ranging from 2-88 Hz in amplitude at 900 MHz and 25°C. The dispersion curves of Val 93 across magnetic field strengths and temperatures are shown in Fig 3.25.a. Millisecond dynamics in the form of relaxation dispersions were widespread across the TF structure as shown in Fig 3.25.c with large amplitude dispersions (large R_{ex}) observed in all TF domains. Exchange contributions to the relaxation rates were particularly large for the RBD, the arm 2 region of the SBD and the residues surrounding the C-terminal helix. Resonances displaying a clear absence of millisecond dynamics mainly map to the interior of the PPiase domain.

The exchange contribution to the relaxation rate (R_{ex}) can be readily extracted from the CPMG RD curves by determining the difference between the highest and lowest values in effective R_2. However, R_{ex} depends nonlinearly on the exchange rate, populations and the chemical shift difference between states, meaning that the amplitude of the R_{ex} does not directly reflect the amplitude or rate of the motion. Therefore, the 8 different experimental conditions were fitted simultaneously in order to extract precise kinetic

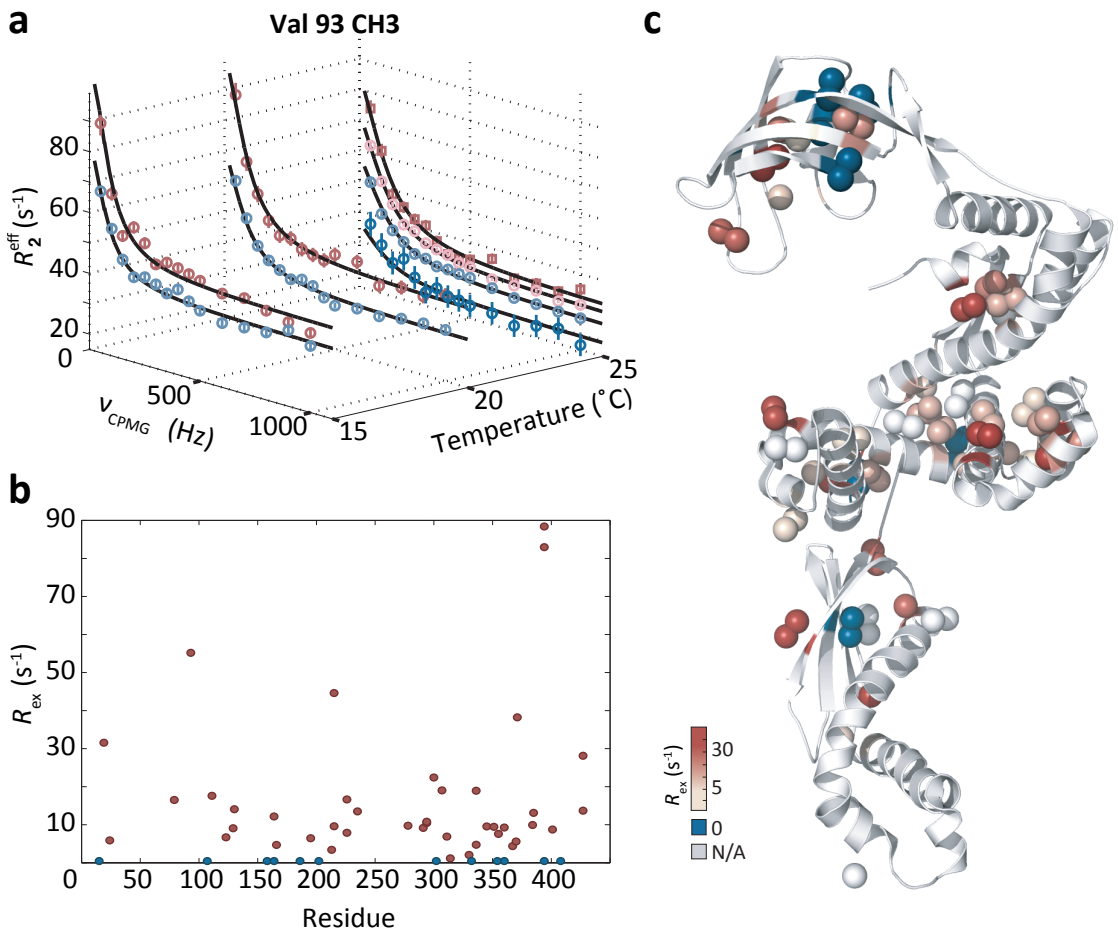

Figure 3.25: Millisecond dynamics in trigger factor. (**a**) CPMG relaxation dispersion data for Val 93. (**b**) Sequence distribution across TF of experimentally determined R_{ex} values at 25°C and 900MHz. (**c**) Structural distribution of R_{ex} values presented in b.

information.

Determination of the exchange rate and population sizes of the observed dynamics : Dispersion curves were fitted to equations 1.1.1-1.1.5, Section 2.1.6.2 in order to extract the values of the exchange rate, population sizes and chemical shift difference using GUARDD. Measurements at multiple magnetic field strengths and temperatures were required for the accurate deconvolution of the various parameters as discussed in Section 2.1.6.2. Although multiple-quantum dispersion measurement were made, in which both ^{1}H and ^{13}C chemical shift differences contribute to the detected dispersion, for their increased sensitivity over single quantum experiments, the data was treated as single-quantum data when visual inspection showed that proton chemical shift difference contributions were negligible. This approach significantly improved the quality of the fits due to the reduced

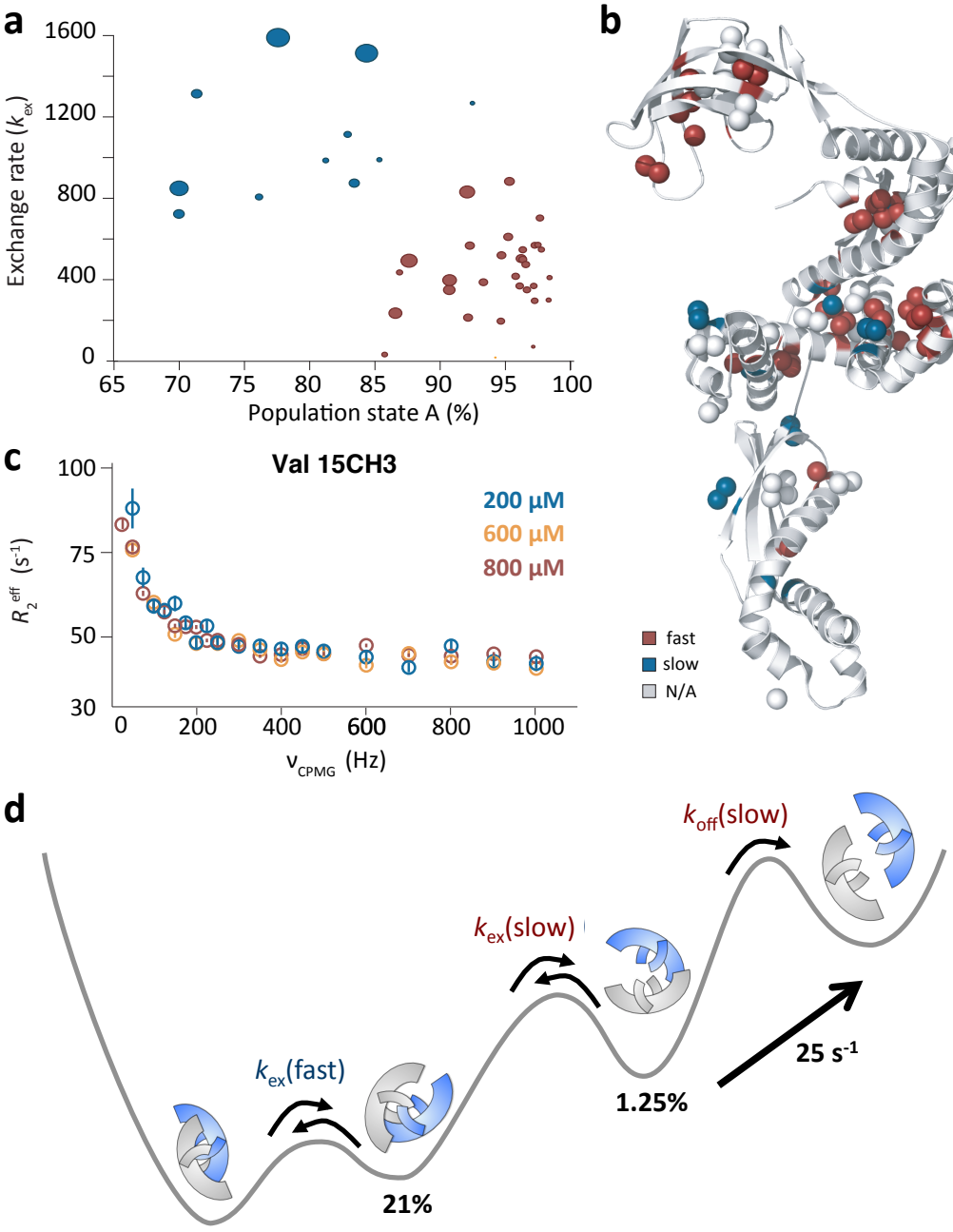

Figure 3.26: Proposed model for the mechanism of dimerisation based on CPMG relaxation dispersion experiments. (**a**) Experimentally derived exchange rates and size of the population of state A based n the CPMG relaxation dispersion curves of individual resonances. Circles size are relative to the amplitude of R_{ex}. Colours correspond to slow (red) and faster (blue) dissociation rates (**b**) Structural distribution of fast (blue) and slow (red) exchanging residues across the TF structure (**c**) CPMG relaxation dispersion curves of residue V15 at 200 (blue), 600 (yellow) and 800 μM (red)TF^{ILV}. (**d**) Preliminary model for the step-wise dissociation of the TF dimer based on the experimentally derived exchange kinetics presented in a.

number of free parameters.

Fitted exchange rates and ground state (state A) populations are plotted in Fig 3.26. A large majority of fitted dispersions cluster to small excited state populations (6%) and relatively slow exchange rate values (425 s^{-1}). Outliers formed a broad cluster around an excited state population of 21% and a faster exchange rate of 1100 $^{-1}$. From this analysis it appears that the majority of residues are involved in an exchange event with a forward rate of 400 s^{-1} and off-rate of 25 s^{-1} .

Model of the step-wise mechanism of dimer dissociation : We hypothesise that origin of the observed widespread millisecond dynamics is the TF dimerisation. The concentration dependence of the dispersions was therefore investigated since R_{ex} is highly sensitive to change in the population distribution. No concentration dependent behaviour was observed (Fig 3.26.c) negating in first instance the dimerisation as origin of the dispersions. Furthermore, the off-rate determined for the exchange is more rapid that has previously been determined (ca. 1 s^{-1}, [148]). However, the majority of residues displaying dispersions were implicated in the dimerisation process in the study outlined Section 3.5.1. It was further noted that two symmetrical exchange events with the same experimentally determined slow exchange rate (425 s^{-1}) gives rise to an theoretical off-rate of 1.6 s^{-1} in close agreement with previously published values. The inclusion of the additional observed fast exchange process, results in a pre-equilibrium of 1.25% excited state and a global dissociation rate 0.43 s^{-1}.

We therefore propose a model for the dissociation of the TF dimer based on the stepwise dissociation of two or three distinct dimer interfaces (Fig 3.26.d). Only the simultaneous dissociation of all sites results in complete monomerisation giving rise to the slower dynamics observed on a macroscopic scale than detected by CPMG relaxation dispersion measurements.

3.6 Concluding remarks

The advantages of the selective methyl labelling strategy detailed in Section 3.3 have allowed the highly detailed study of the ∼100 kDa TF dimer. TF^{ILV} has proven suitable to detailed structural investigation because of relative high stability at 25°C (∼ 4 days) and compatibility with physiological conditions used for ribosome studies (pH 7.5 and temperatures ranging from 4-25°C). The availability of a partial assignment of side-chain methyl group resonances means that conformational and dynamic changes could be monitored across the TF sequence.

The extensive prior studies into the various TF equilibria permits the mathematical modelling of the complex TF equilibrium. This provides a valuable tool in the predic-

tion of population distribution under specific experimental conditions but this ensemble study of the various equilibria further sheds light on the biological implications of the competition between the various reactions. The TF equilibrium appears highly fine-tuned to its function as non-selective molecular chaperone at the ribosomal exit tunnel, while providing a buffer in case of aggregation-prone protein accumulation under stress.

The combination of intensity analysis, cross-saturation and PRE experiments resulted in a preliminary model of the solution structure of the TF dimer, with a head-to-toe symmetrical arrangement (Fig 3.22). The dimerisation interface was found to involve a large number of contacts and regions identified to be in proximity of the nascent chain and isolated protein substrates (Fig 1.6). Possible competition for binding sites and/or steric hinderance might prevent simultaneous dimerisation and substrate binding.

Pico to nano second dynamics were largely absent in the dimeric TF indicating the main structural features of TF are rigid on this timescale. Small fluctuations in the orientation of secondary structural element relative to each other might exist based on the low order parameters determined for a few key interface residues.

While largely rigid on fast timescales, TF displays significant millisecond timescale dynamics. The main exchange event identified by CPMG relaxation dispersion occurs with an exchange rate of 425 s^{-1} and the minor conformation is populated to 6%. We propose a model for the step-wise dissociation of the TF based on two symmetrical dimerisation interfaces resulting in a cumulative off-rate of 1.6 s^{-1} in close agreement with previously determined kinetics of the dimerisation reaction [148]. The concept of two main dimerisation interfaces is further in agreement with our proposed solution structure of the TF dimer (Fig 3.22).

Careful analysis of the concentration dependence of TF^{ILV} spectra has permitted the determination of a dissociation constant of 7-10μM under our experimental conditions (Fig 3.17). The slight decrease in affinity observed with respect to published K_d values could be explained by the increased viscosity or weakening of hydrophobic interactions in D_2O relative to H_2O. The identification of the individual contribution of the TF dimer and monomer to the ^{1}H-^{13}C HMQC spectrum highlights the significant differences between the two species and suggests the structural investigation of the TF monomer and changes therein in the presence of substrate might be possible despite the impossibility of obtaining a pure monomeric sample under the experimental conditions required for NMR (>100 μM).

Chapter 4

NMR investigations of the interaction of trigger factor with ribosome-nascent chain substrates - α-synuclein

4.1 Introduction

In Chapter 3, the structural and dynamic behaviour of trigger factor (TF) in the absence of substrates was described. This detailed characterisation of the NMR behaviour of TF forms the basis for the interpretation of changes observed in the presence of heterologous substrates. The following parts of this thesis, Chapters 4 and 5, aim to characterise the interaction of TF with protein substrates, on and off the ribosome. NMR spectroscopy is used to provide residue specific details of these interactions.

4.1.1 The TF interaction with protein substrates

As was discussed, Sections 1.2.2.3 and 3.4.1, the TF chaperone appears to promiscuously interact with a large subsection of the *E. coli* proteome and was found to show a particular preference for larger, hydrophobic proteins such as the OMP family of outer membrane proteins [149]. No TF consensus sequence has to date been identified but TF interaction sites appear to be enriched in aromatic and non-polar residues [159]. The analysis of substrate protein sequences found regions of high TF affinity to be correlated with high local hydrophobicity [159]. This has permitted the correct prediction of TF interactions sites in substrates such as isocitrate dehydrogenase (ICDH), the β sub-unit of RNA polymerase (RpoB) and alkaline phosphatase precursor (PhoA) [130, 161]. The interaction of TF

with isolated proteins is typically weak (>100 μM, see Table 3.3) although notable exceptions include a subset of ribosomal proteins [84]. TF interacts significantly stronger with ribosome-associated nascent polypeptides with experimentally determined affinities ranging from 1 nM to an upper boundary of 1 μM: the TF affinity for unoccupied ribosomes is summarised in Table 3.1, Section 3.4.1. The affinity of TF for nascent chains displays both a length- as well as sequence-dependence. *In vivo*, TF preferentially interacts with substrates of over 100 amino acids [149], probably due to competition with other ribosome associated factors, which have been shown in the case of SRP to be in direct competition with TF for binding at ribosomal protein L23 [147, 156].

In vitro, TF affinity correlates with the exposure of hydrophobic, high affinity motifs to the cytosol as they emerge from the ribosomal exit tunnel [148, 161, 240]. The residency time of TF on ribosomes was found to be between 10-15 s, shown to extend up to 35 s in the presence of a high affinity substrate nascent chains by FRET [148, 161]. However, a recent study found the kinetics of ribosome association to be more rapid than this, with half-lives of 0.06 s and 2 s determined for the ribosome and RNC interactions, respectively [162] and noted that the previous results are likely to have been artefacts of the specific experimental method. Residency of TF on the ribosome as measured by half-lives and off-rates correlates with an increase in detected affinity for a specific nascent chain substrate, which indicates that the presence of a high affinity motif does not affect the initial recruitment of TF to the ribosome but results in the stabilisation of the TF-ribosome complex.

The substrate binding sites of TF have been described in detail in Section 1.2.2.3. An important recent NMR study of the TF interaction with unfolded PhoA identified four distinct substrate interaction sites each with specific amino acid composition that appear to result in a differential preference for certain types of substrates [130]. The PPIase domain in particular prefers clusters of aromatic residues while sites in the SBD mainly interact with non aromatic, hydrophobic residues although it can further accommodate peripheral electrostatic interactions. On the ribosome, nascent chains were additionally shown to be in proximity of the RBD which exposes a more hydrophobic surface when bound to the ribosome [131, 132].

4.1.1.1 α-synuclein

In this Chapter, α-synuclein is used as a model protein for the interaction of TF with protein substrates. α-synuclein is an intrinsically disordered protein (IDP) whose function is generally not well understood but may have a role in lipid transport via SNARE complexes [246] and has been implicated in Parkinson's disease [247, 248]. α-synuclein self-associates into amyloid fibrils leading to the formation of large protein aggregate deposits known as Lewy bodies in Parkinson's patients [249]. Due to its dynamic nature, α-synuclein

and its conformational behaviour has been extensively studied by NMR spectroscopy and has been amongst the first IDPs where preferred conformations in the absence of rigid secondary structure were described [250, 251].

The α-synuclein primary sequence is typically subdivided into three regions (Fig 4.1 and Appendix A.2). The first 100 N-terminal amino acids form an amphipathic region with helical propensity. In the presence of lipid, this region forms two long α-helices that interact with the lipid surface [252, 253]. The non-β-amyloid component (NAC) domain (61-95) constitutes the most hydrophobic region of the protein sequence and is essential for self polymerisation of α-synuclein [254]. The 40 C-terminal residues carry a large number of negatively charged residues and appear devoid of any propensity to form secondary structure.

The α-synuclein ribosome nascent chain complex (RNC) has been previously used in studies of the interaction of TF with substrate proteins where it was shown that the presence of the α-synuclein nascent chain does not affect the affinity of TF for ribosomes [144]. Cross-linking studies did however demonstrate that TF is in proximity of the α-synuclein nascent chain at the exit tunnel [153]. Here, it does not convey protease protection to the α-synuclein nascent chain suggesting α-synuclein maintains significant conformational freedom in the presence of TF [153]. Using these studies as a basis, the α-synuclein RNC was considered a suitable null model for TF interaction with nascent chains at the ribosomal exit tunnel.

4.2 Methods

Preparation of isolated α-synuclein protein and mutants followed protocols described in detail in Section 2.2.4.4 and TF was prepared as outlined in Section 2.2.4.2. The NMR experiments and experimental conditions are described in Section 2.3. Preparation of 70S ribosomes and α-synuclein RNC samples are detailed in Section 2.2.5 and 2.2.6 respectively. The mathematical models employed in Section 4.6 are described in greater detail in this Section and the relevant Matlab scripts have been included in Appendix C.

4.3 Strategy

The absence of structure in the native α-synuclein protein permits the study of the TF interaction with ribosome-nascent chain complexes (RNCs) unhindered by additional complexity in the form of co-translational folding of the substrate protein. The α-synuclein RNC has been the topic of extensive development within our laboratory and shown to produce a well resolved NMR spectrum under favourable experimental conditions of low temperatures and physiological pH (Weise *et al.*, in preparation). The absence of an appar-

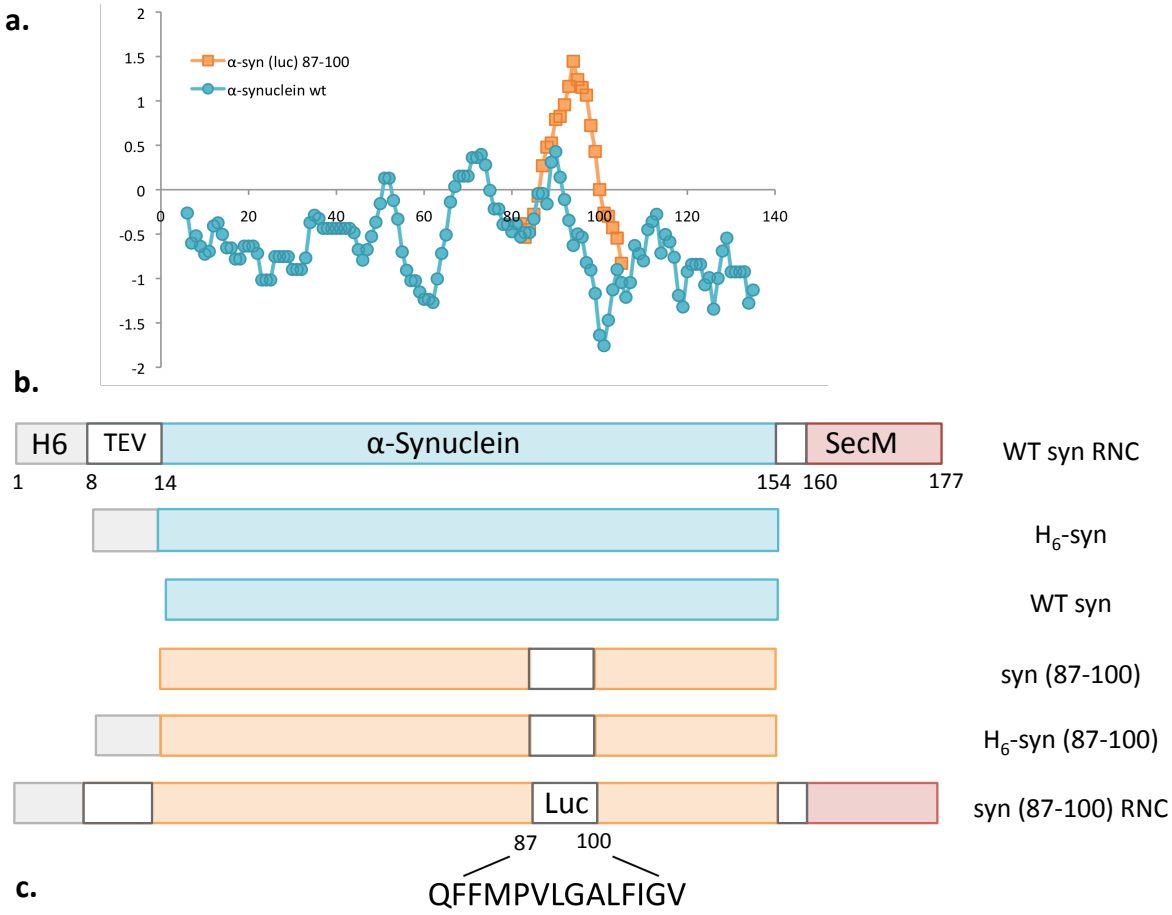

Figure 4.1: α-Syn(luc) chimera RNC constructs used in this study. (**a**) Roseman hydrophobicity plot of the developed RNC constructs colour coded according to b . (**b**) The various protein and RNC constructs used in this study. (**c**) The firefly luciferase sequence substituted in to α-syn (Luc 87-100) RNC is shown below the relevant region.

ent, specific TF binding site suggests the α-synuclein RNC may maintain enough flexibility on the ribosome in the presence of TF to remain observable by NMR spectroscopy, in a similar fashion to the L7/L12 stalk region (Section 1.1.3.1).

The α-synuclein RNC was produced following a strategy developed by Annika Weise (UCL) by the addition of the 17 residue SecM translational stalling sequence at the C-terminus to the H_6-synuclein amino acid sequence (Fig 4.1). A TEV protease site was included to permit the removal of the H_6-tag after purification of the RNC. H_6-α-synuclein and WT α-synuclein were used as isolated reference proteins in order to characterise the TF interaction with α-synuclein in the absence of the ribosome.

An additional model substrate was designed based on the wild-type α-synuclein RNC sequence. This approach extends previous work by Lakshmipathy *et al.* [144] using the substitution of a known, high affinity sequence from firefly luciferase to enhance the TF

affinity for α-synuclein. Substitution of residues 87-100 of luciferase into the same position in α-synuclein replaces a sequence of mixed hydrophobic and hydrophilic nature for a highly hydrophobic one with proven affinity for TF. The inclusion of the firefly luciferase sequence was shown, using fluorescence, to enhance the recruitment of TF to the α-synuclein RNC [144]. The inclusion of a high affinity site into the α-synuclein RNC was undertaken to allow the extension of the present NMR-based study from the characterisation of non-specific interactions of TF with substrates at the ribosomal exit tunnel to include specific, targeted interactions with emerging nascent chains. An overview of TF substrate constructs used in this study is presented in Fig 4.1.

4.4 Interactions of TF with isolated protein substrates

Previous studies of the TF interaction with substrate proteins have focussed primarily on the interaction at the ribosomal exit tunnel with emerging nascent chains. In the current study, the interaction of TF with various α-synuclein substrate proteins in the absence of ribosomes is in the first instance investigated in order to avoid the complexity of multiple TF states in the analysis of the NMR data (Section 1.2.2.5 presents an overview of these potentially complicating TF equilibria) and avoids experimental constraints associated with the ribosome such as sample stability and limited maximum achievable 70S particle concentrations, thereby permitting a broader range of NMR techniques. Further complexity will be added towards the end of this chapter where the nascent chain equivalents of the here described substrate proteins are studied in the presence of TF.

4.4.1 Characterisation of the interaction of TF with WT α-synuclein

The ^{1}H-^{15}N HSQC fingerprint spectrum of 100 μM ^{15}N WT α-synuclein in Tico buffer (H_2O) pH 7.5 at 4°C is well resolved although little dispersion in the ^{1}H dimension is observed (all signals between 7.5-8.5 ppm) characteristic of an unfolded protein (Fig 4.2). The addition of a N-terminal H_6-tag to the α-synuclein sequence does not significantly change the ^{1}H-^{15}N HSQC spectrum of 100 μM α-synuclein at 4°C and pH 7.5. Chemical shift changes of resonances originating from residues in close proximity (up to 4 residues removed) were detected (average 0.05 ppm in ^{1}H dimension and 0.4 ppm in ^{15}N dimension) but the combination of a previous assignment of NMR resonances of the His-tagged α-synuclein at pH 6.5 and the WT α-synuclein at pH 7.5 in our laboratory [255, 256] allowed for ready assignment of the H_6-synuclein spectrum (Fig 4.2). Resonance intensities mapping to the first 30 residues were however, seemingly randomly, perturbed suggesting that the presence of the H_6-tag modifies the properties of the α-synuclein N-terminus.

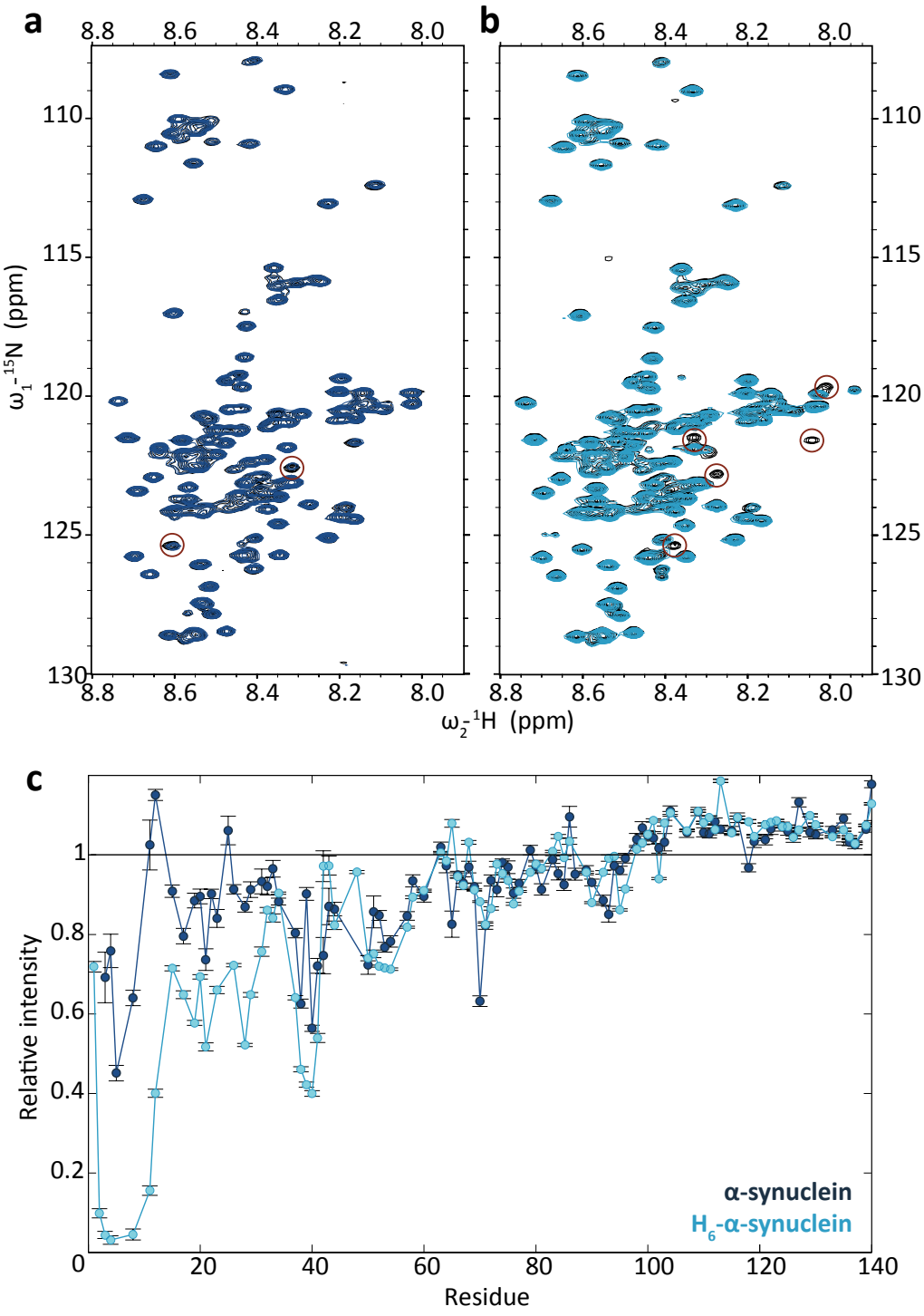

Figure 4.2: TF interaction with isolated α-synuclein and the effect of the H_6 purification tag. (**a**) ^{1}H-^{15}N HSQC overlay of 100 μM WT α-synuclein before (black) and after equimolar addition of TF (blue) at 4°C. (**b**) ^{1}H-^{15}N HSQC overlay of 100 μM H_6-synuclein before (black) and after equimolar addition of TF (cyan) at 4°C. Areas of significant broadening are circled in red. (**c**) Plot of relative ^{1}H-^{15}N HSQC resonance intensities upon addition of TF to α-synuclein and H_6-α-synuclein. Errors were determined from the standard deviation of 100 resonances in the noise of the NMR spectra.

Interaction of TF with WT α-synuclein: Addition of equimolar H_6-TF results in minor broadening of specific ^{1}H-^{15}N HSQC resonances. The detected broadening is strongest (down to 70%) at the extreme N-terminal residues of α-synuclein and affects residues up to Lys 60, as shown in Fig4.2.c. The observed changes in resonance intensities are an indicator of the interaction of TF with WT α-synuclein around the affected residues. The interaction results in the transient increase of the apparent α-synuclein size from 14 kDa to 62 kDa (or 110 kDa dependent on whether the interaction occurs with monomeric or dimeric TF). This transient increase in size and the exchange between the two states, free and bound, leads to increased linewidths and reduced peak heights. Section 2.1.2 discusses the effects of molecular size and chemical exchange on the NMR signal.

From the observed resonance intensity changes, TF appears to interact with α-synuclein at the extreme N-terminus and around residue Tyr 39 where the broadening is most pronounced, to ~50 % the resonance intensity in the absence of TF. Both interactions are centred around aromatic residues.

Interaction of TF with H_6-α-synuclein: Addition of equimolar H_6-TF to H_6-α-synuclein lead to significantly stronger broadening events at the α-synuclein N-terminus. Residues 1-6 (amino acid sequence in Appendix A.2) are in fact broadened beyond detection as shown in Fig 4.2.c. The observed resonance broadening centres around residue Phe 4 with a possible second site of interaction detected at residue Tyr 39 where resonances are broadened to 40% of the original intensity. The detected broadening extends along the sequence to residue Lys60 where peak intensities are no longer perturbed.

Discussion: The sensitivity of NMR has permitted the detection of a transient interaction at the α-synuclein N-terminus around aromatic Phe 4. The presence of a hexahistidine purification tag appears to significantly enhance the affinity of TF for α-synuclein. The broadening detected at the α-synuclein N-terminus is doubled in the presence of the H_6-tag. The interaction at Tyr 39 further appears to be enhanced slightly. The residues linking the two identified sites of interaction are also more significantly affected by the presence of TF in his-tagged α-synuclein. Since the chemical environment at residues beyond Ala 30 is unaltered by the addition of the H_6-tag in this natively disordered protein, it is likely that the increased local concentration due to the heightened affinity at the N-terminus, enhances the interaction with TF at the second binding site instead of a direct effect from the H_6-tag.

The interaction with TF appears driven by key aromatic residues (Phe 4 and Tyr 39) at the N-terminus of the α-synuclein correlating with previous reports on the nature of the TF interaction with substrate proteins (Section 1.2.2.3). The precise mechanism through which the H_6-tag increases recruitment of TF is unclear in particular since the his-tag does

not appear to be involved in the actual interaction as shown by the limited broadening observed for residue Ser 0. It is thus likely that the increased affinity of TF is related to the altered properties of the first 30 residues upon inclusion of a H_6 purification tag. The transient interaction of TF at the N-terminus in the presence of the H_6-tag effectively alters the chemical environment and/or conformational space of 50%of the sequence of this natively unfolded, highly dynamic protein.

4.4.2 Characterisation of the interaction of TF with α-synuclein (Luc 87-100)

4.4.2.1 Expression and purification of α-synuclein (Luc 87-100) mutants

The α-synuclein substitution mutant which includes firefly luciferase residues 87 to 100 (α-syn (Luc 87-100), Fig 4.1) was successfully heterologously expressed and purified from *E. coli* as described in Section 2.2.4.4. The full protein sequence can be found Appendix A.1.2. Purification followed a similar protocol as described for WT α-synuclein.

4.4.2.2 Assignment of the ^{1}H-^{15}N HSQC spectrum of H_6-syn(Luc 87-100)

The ^{1}H-^{15}N HSQC spectrum of H_6-α-syn (Luc 87-100) is very similar in appearance to the H_6-α-synuclein spectrum (Fig 4.4). α-synuclein residues up to three amino acids removed from the site of substitution do not show significant changes between the two spectra. Residues closer to the substitution show small changes in chemical shift positions (average 0.03 ppm in ^{1}H dimension and 0.15 ppm in ^{15}N dimension) while mutated residue resonances are clearly absent and a number of new resonances (13) have appeared corresponding to the new amino acid sequence.

In order to study the interaction of TF with H_6-α-syn(Luc 87-100) at a residue specific level, ^{1}H-^{15}N HSQC resonances originating from the substituted residues were assigned using triple resonance experiments. Uniformly ^{13}C/^{15}N labelled protein samples were prepared for an assignment strategy based on amino acid type identification from the HNCACB/HNcoCACB set of NMR experiments and identification of sequential resonances from the HNCO and HNcaCO pair of 3D experiments shown in Fig 4.3. The same NMR data were used to validate the transferred assignment of non-mutated residues from the WT α-synuclein HSQC spectrum. All backbone amide group resonances were successfully assigned as well as all CO, Cβ and 99% of Cα resonances (Appendix A.5).

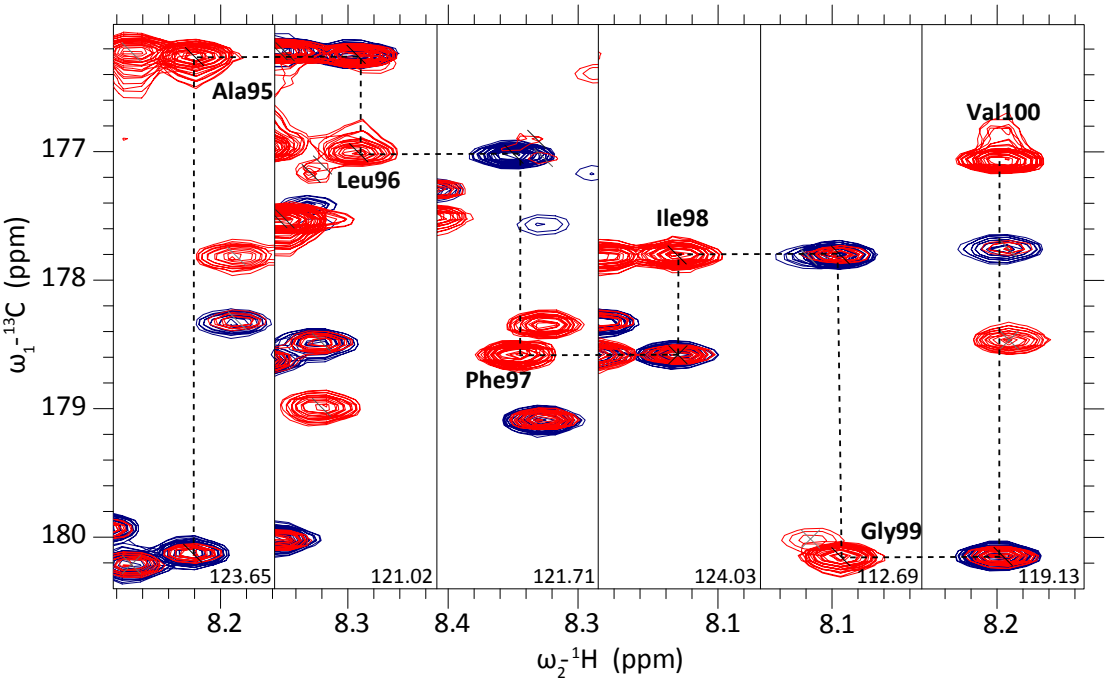

Figure 4.3: NMR backbone assignment of α-syn (Luc 87-100). Representative overlays of ^{1}H-^{13}C strips of the HNCO (blue) and HNcaCO (red) 3D experiments displaying the backbone connectivity determined for the substituted residues 95-100.

4.4.2.3 TF interaction with α-synuclein (Luc 87-100)

Addition of equimolar H$_6$-TF^{ILV} to 100μM ^{15}N H$_6$-syn(Luc 87-100) at 4°C and pH 7.5, resulted in the selective broadening of residues in four distinct regions. Resonance broadening was observed around residues Phe 4, Tyr 39, His 50 and the inserted region, in particular residues 89-100. Resonances originating from the luciferase substitution (89-100) are all broadened beyond detection at equimolar concentrations of H$_6$-TF^{ILV} to H$_6$-syn(Luc 87-100).

From the highly broadened resonances originating from residues in the luciferase substitution site (87-100), the resonance intensities increase steadily towards the N-terminus until the preceding site of interaction at His 50 is encountered. The sequence C-terminal to the substitution site, on the contrary, rapidly recovers its associated resonance intensity to that in the absence of TF. The substitution of a 14 amino acids firefly luciferase sequence at positions 87-100 effectively introduces a site for specific TF interaction. In combination with the existing N-terminal interaction in the WT α-synuclein sequence, the new TF interaction site at 89-100 appears to tether the α-synuclein protein and restrict the conformational freedom of the first 100 residues of α-synuclein.

Moreover, this 'tethering' effect appears to enhance the transient interactions at Tyr

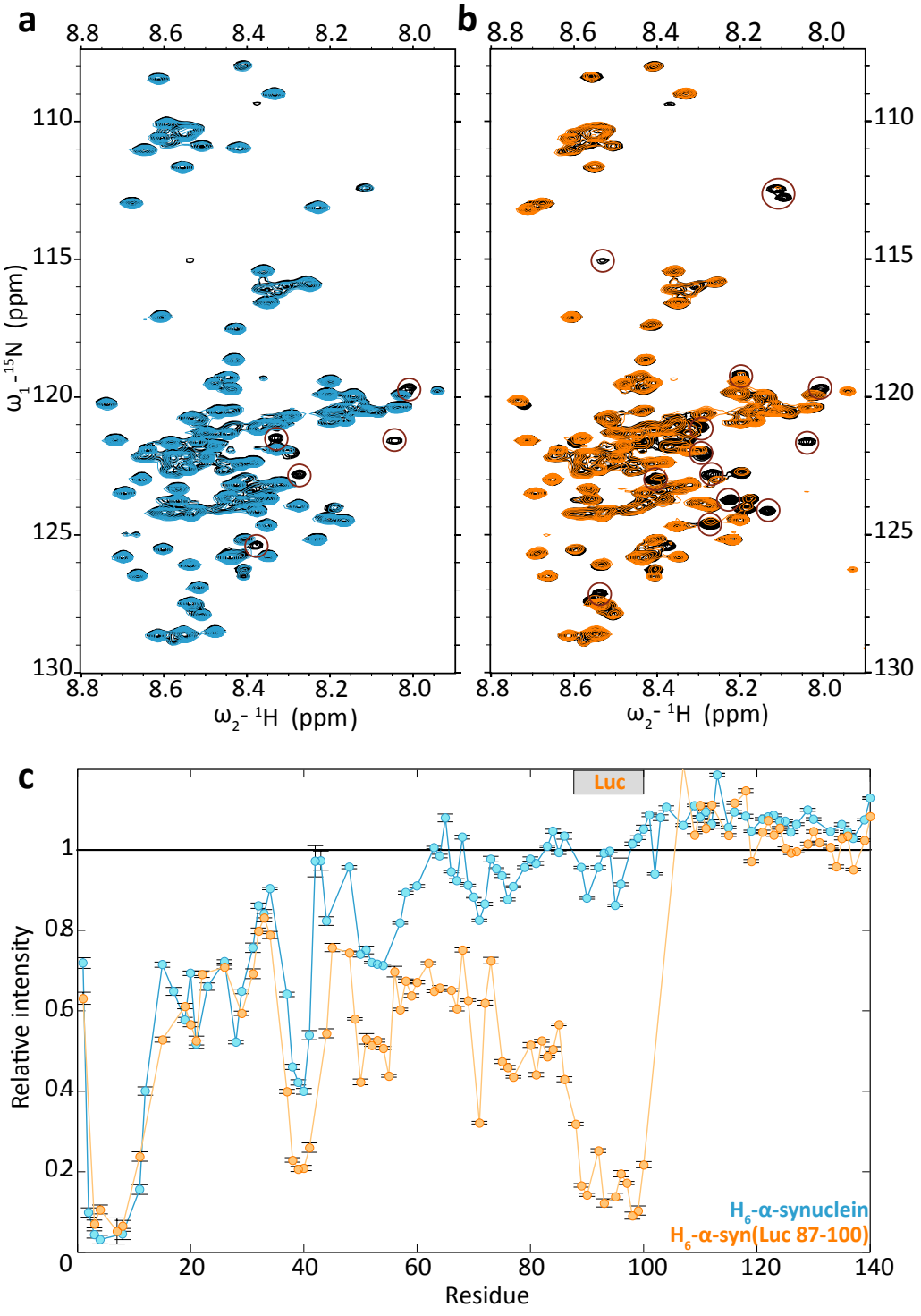

Figure 4.4: TF interaction with isolated α-synuclein (Luc 87-100). (**a**) ^{1}H-^{15}N HSQC overlay of 100 μM H$_6$-α-synuclein before (black) and after equimolar addition of TF (cyan) at 4°C. (**b**) ^{1}H-^{15}N HSQC overlay of 100 μM H$_6$-αsyn (Luc 87-100) before (black) and after equimolar addition of TF (orange) at 4°C. Areas of significant broadening are circled in red. (**c**) Plot of relative ^{1}H-^{15}N HSQC resonance intensities upon addition of TF to α-synuclein and H$_6$-α-synuclein. Errors were determined from the standard deviation of 100 resonances in the noise of the NMR spectra.

39 and His 50 with resonance broadening stronger for H_6-α-syn(Luc 87-100) than in WT H_6-α-synuclein upon addition of TF. Finally, the N-terminal interaction itself appears unperturbed by the presence of a second higher affinity binding site suggesting the two binding events do not compete. This could indicate the presence of multiple substrate binding sites on TF with differential binding properties.

Titration of H_6-α-syn(Luc 87-100) with H_6-TF^{ILV}: Gradual titration of 100 μM H_6-α-syn(Luc 87-100) with H_6-TF^{ILV} was undertaken at 25°C and pH 6.5 in order to allow the simultaneous monitoring of both H_6-α-syn(Luc 87-100) ^{1}H-^{15}N HSQC and H_6-TF^{ILV} ^{1}H-^{13}C methyl TROSY HMQC resonances over the course of the titration. Due to the unstructured nature of α-synuclein, the rapid solvent exchange of amide protons leads to significant broadening of resonances in the α-synuclein ^{1}H-^{15}N HSQC spectrum under the experimental conditions used thus far in the study of TF by NMR (25°C and pH 7.5).

Titration of ^{15}N H_6-α-syn(Luc 87-100) with H_6-TF^{ILV} at pH 6.5 and 25°C permits the identification of three distinct sites of broadening at sub-stoichiometric TF concentrations as shown in Fig 4.5.a. The observed broadening is strongest for the luciferase substituted region, residues 89 to 100. Residues most strongly affected within this region are broadened beyond detection at 0.75 equivalence TF while resonances less affected by broadening (i.e Gly99) show that the interaction at this site saturates between 1 and 1.5 molar equivalence of H_6-TF^{ILV} (Fig 4.5.c).

The interaction at the extreme N-terminus of the α-synuclein sequence around Phe 4, appears to saturate at higher concentrations of TF, between 1.5 and 2 molar equivalence of H_6-TF^{ILV} (Fig 4.5.c). The final interaction site at His 50 does not appear to saturate at the concentrations sampled in this titration. The relative strengths of the interactions can thus be tentatively ranked as 89-100 > Phe 4 > His 50. The numerical fit of the titration curves for residues 8 and 100 where resonances can be observed across the entirety of the titration produce dissociation constants of 4μM and 1μM for the interactions at the N-terminus and the substituted region (87-100) respectively. Concentrations within the range of the dissociation constants are not adequately sampled over the course of this titration resulting in inaccuracies of the determined dissociation constants; their precise values are likely to be in this low μM range however.

At stoichiometric concentrations of added TF, additional sites of broadening are observed around Tyr 39 and Val 70 which do not saturate over the course of this titration (Fig 4.5.b). In summary, at 25°C and pH 6.5, TF appears to preferentially interact with aromatic and hydrophobic residues in H_6-α-synuclein (Luc 87-100) with interactions centred around Phe and His and Tyr residues. The observed ranking of the affinities of individual regions of α-synuclein (Luc 87-100) suggest that TF interaction decreases with the polarity of the aromatic residue.

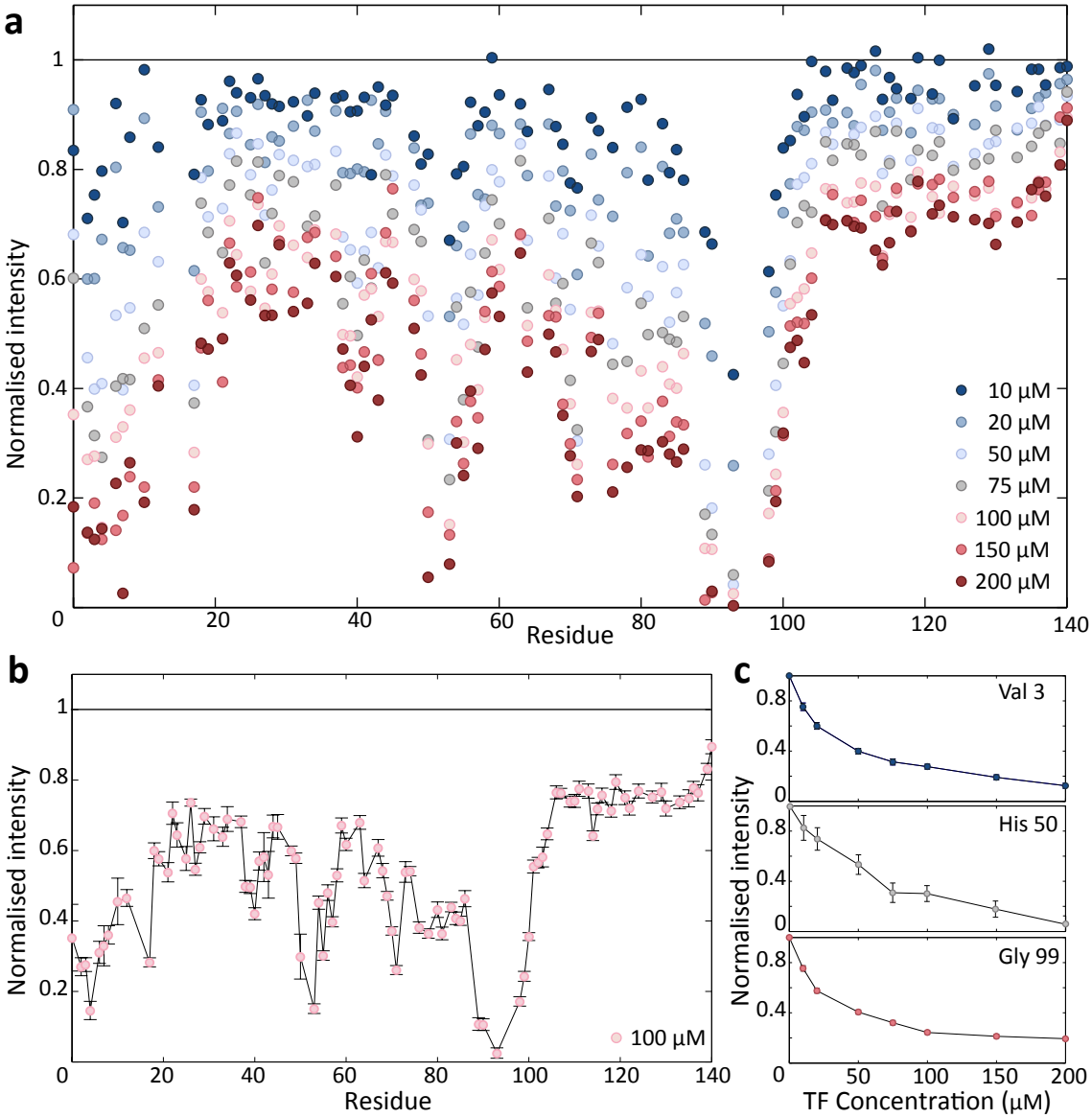

Figure 4.5: TF interaction with isolated α-synuclein (Luc 87-100). (**a**) Plot of relative ^{1}H-^{15}N HSQC resonance intensities upon gradual addition of 10, 20, 50 ,75 ,100 ,150 and 200 μM TF to 100 μM H_6-α-syn (Luc 87-100). (**b**) Relative intensities at equimolar H_6-TF and α-syn (Luc 87-100). Errors were determined from the standard deviation of 100 resonances in the noise of the NMR spectra (**c**) Relative intensities of selected α-syn (Luc 87-100) residues over the course of the TF titrations.

4.4.2.4 Temperature and pH dependence of the TF interaction with α-synuclein (Luc 87-100)

The TF interaction with H_6-α-synuclein (Luc 87-100) was monitored at different temperatures (4°C and 25°C) and pH conditions within the physiological range (6.5 and 7.5) at equimolar concentrations (100μM) by ^{1}H-^{15}N HSQC experiments. The main contact sites described above are conserved across all experimental conditions, although variations in the extent of broadening are observed for specific sites shown in Fig 4.6. The N-terminal interaction appears temperature sensitive with a greater extent of broadening detected at lower temperatures (from 30% at 25°C to 10% at 4°C of the resonance intensity in the absence of TF), affecting residues over a longer stretch of the protein sequence (residues 1-10). The interaction at His 50 also displays temperature dependence but appears stronger at higher temperatures (40% and 20% of the resonance intensity in the absence of TF at 4°C and 25°C respectively).

The interaction at Tyr 39 and the behaviour of the α-synuclein (Luc 87-100) C-terminus are affected by changes in the pH. The interaction at Tyr 39 is enhanced slightly (from

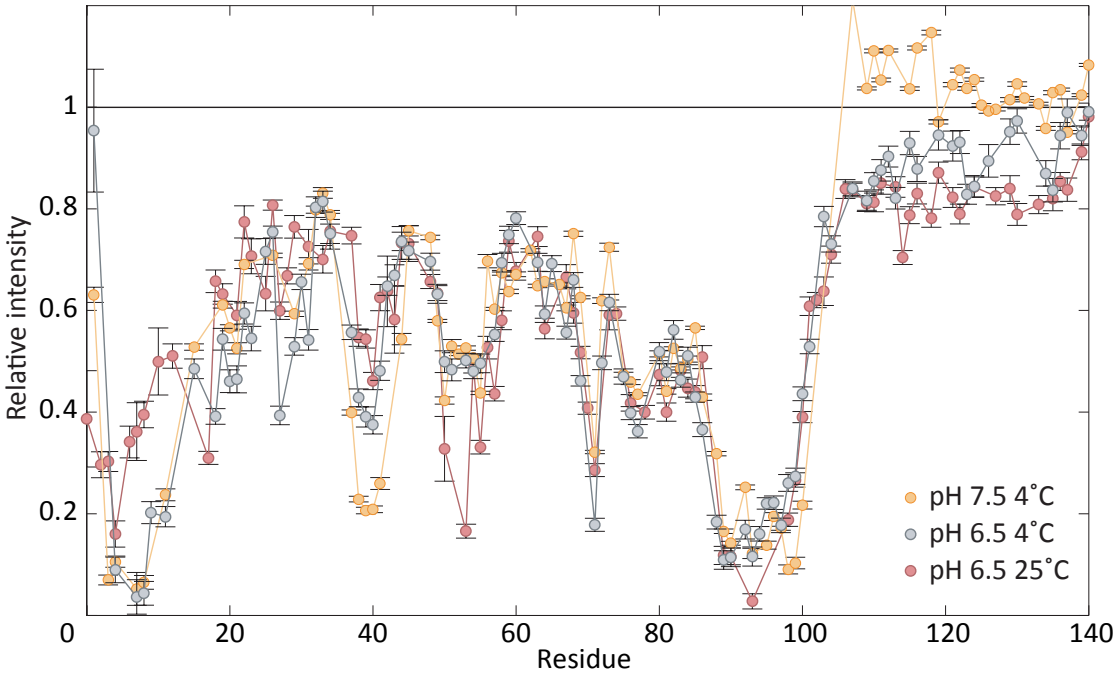

Figure 4.6: pH and temperature dependence of TF interaction with α-synuclein (Luc 87-100). Relative ^{1}H-^{13}C HSQC resonance intensities upon addition of equimolar H_6-TF to 100μM α-synuclein (Luc 87-100) under various experimental conditions: 4°C and pH 7.5 (orange), 4°C and pH 6.5 (grey) and 25°C and pH 6.5 (red). Errors were determined from the standard deviation of 100 resonances in the noise of the NMR spectra.

40% to 20% of the resonance intensity in the absence of TF) upon increase to 7.5 of the pH. The increased intensities at the C-terminus at higher pH (from 80% to 100% initial signal intensity) could be related to the abundance of glutamate and aspartate residues in this region of the α-synuclein (Luc 87-100) protein. No changes at the 89-100 interaction sites are observed. Resonances originating from residues in this region are however (nearly) broadened beyond detection such that fluctuations in the signal might be difficult to be observed.

Overall, the behaviour of the various identified interaction sites across experimental conditions provides some initial insights into the nature of the different binding events. The temperature sensitivity of some sites suggests hydrophobicity or conformational flexibility might play an important role in mediating the interaction while the pH dependence of other sites suggest electrostatics mediate the TF interaction with H_6-α-syn (Luc 87-100). The differential behaviour of the various sites suggests, 1) that the interactions are independent and do not reflect a single binding event but multiple individual interactions and 2) that the TF interaction with substrate proteins is highly adaptive, capable of accommodating substrates of mixed chemical properties.

4.4.3 TF interaction with α-synuclein substrates by ^{13}C NMR

In order to monitor the changes to the TF conformation upon interaction with the above discussed α-synuclein substrates, the titrations were undertaken using perdeuterated, selective protonated TF^{ILV} to permit simultaneous detection of the behaviour of the α-synuclein substrate and TF. The interaction was monitored at 25°C where the TF ^{1}H-^{13}C HMQC spectrum can be readily detected (as discussed Section 3.3) even under protonated (H_2O based) buffer conditions.

Fig 4.7 shows the H_6-TF^{ILV} ^{1}H-^{13}C HMQC spectrum at pH 7.5 upon addition of equimolar H_6-α-synuclein (blue) and H_6-syn(Luc 87-100) (orange). While small differences were observed in a limited number of resonances in the presence of H_6-α-synuclein, significantly larger perturbations were detected in the H_6-TF^{ILV} ^{1}H-^{13}C HMQC spectrum in the presence of H_6-α-syn(Luc 87-100). Changes to the ^{1}H-^{13}C HMQC are manifested in the form of chemical shift changes and selective broadening of resonances. Resonance intensity reductions of up to 40% in the intensity in the presence of H_6-α-syn(Luc 87-100) map to residues in the RBD and arm regions of the SBD. Chemical shift perturbations were of the order of 0.01 ppm in the ^{1}H dimension and their behaviour across the titration undertaken at pH 6.5 with H_6-α-synuclein (Luc 87-100) indicate the observed phenomenon is in the slow exchange regime.

At excess concentrations of H_6-syn(Luc 87-100), as illustrated in Fig 4.7 in pink at 800 MHz, the observed changes become further pronounced. While the perturbations are wide-

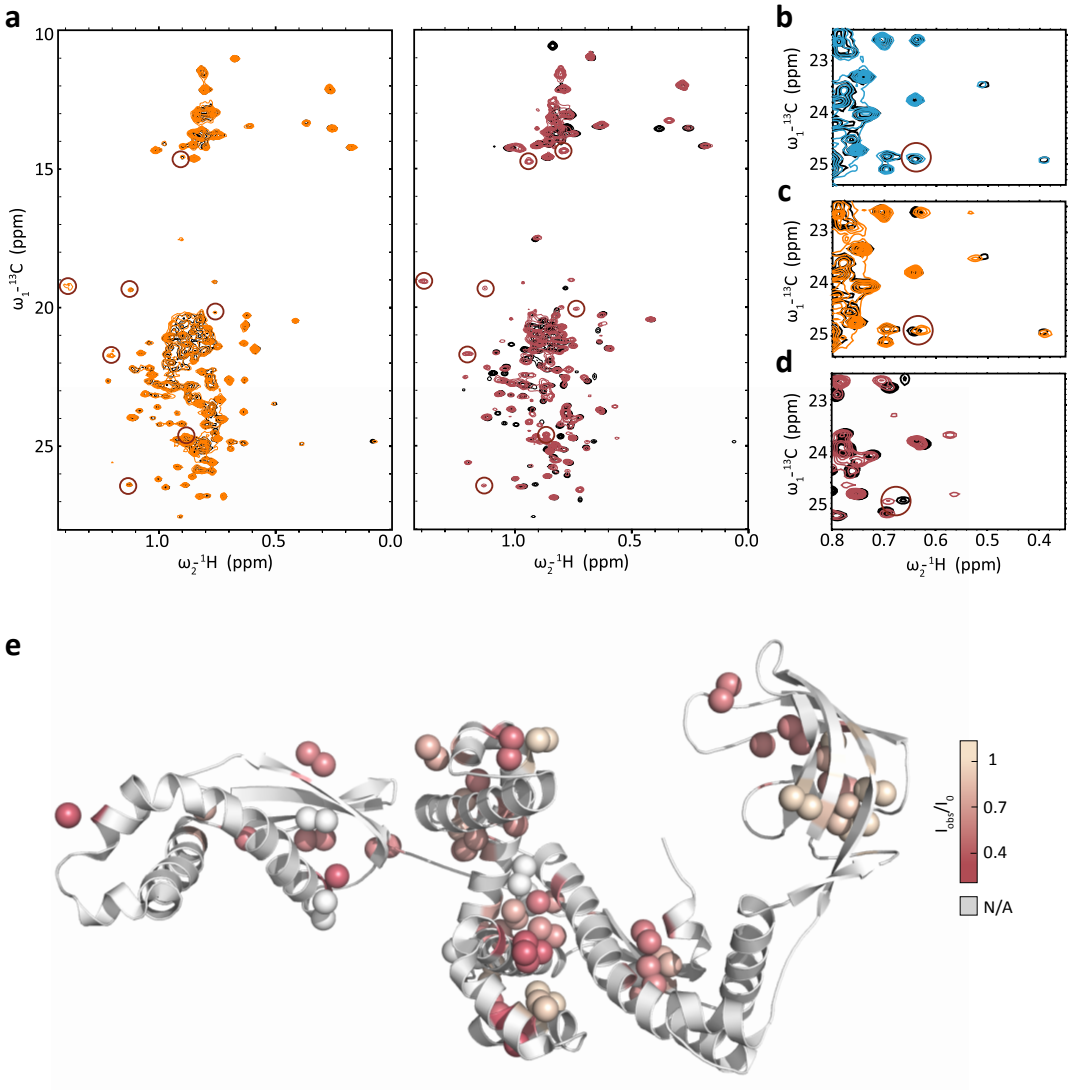

Figure 4.7: Effects of substrate interaction on TF. (**a**) ^{1}H-^{13}C methyl-TROSY HMQC spectra of H_6-TF and TF^{ILV} before (black) and after addition of 2.5 fold excess H_6-α-syn(Luc 87-100) (orange and pink). Novel resonances, not observed in the absence of substrate are circled in red. (**b**) Region of the^{1}H-^{13}C methyl-TROSY HMQC spectra of 100μM H_6-TF^{ILV} before (black) and after addition of equimolar H_6-α-synuclein (blue) (**c**) and H_6-α-syn(Luc 87-100) (orange) as shown in a. (**d**) Region of the^{1}H-^{13}C methyl-TROSY HMQC spectra of TF^{ILV} before (black) and after addition of 2.5 fold excess H_6-α-syn(Luc 87-100) as in a. Areas of significant spectral changes are circled in red.(**e**) Structural distribution of the relative resonance intensities in the absence and presence of 2.5 fold excess H_6-α-syn(Luc 87-100).

spread, the changes in chemical shifts are relatively small (0.01-0.02 ppm) indicating no dramatic changes in TF structure occur upon interaction. The TF dimer ^{1}H-^{13}C HMQC resonances are significantly broadened to 10% reflecting the change in resonance positions. The structural distribution of the intensity changes is shown in Fig 4.7.e. Two main TF areas are affected by substrate interaction, the RBD and arm 2 of the SBD. The extremity of arm1, the neck region and a part of the PPIase domain are further affected indicating the TF interaction with the H_6-syn(Luc 87-100) substrate results in perturbations to all TF domains. A significant number of novel resonances are further detected which do not have a clear origin in the isolated TF^{ILV} spectrum.

Pairs of resonances in slow exchange at 2.5x fold excess of substrate were analysed and their signal integrals extracted. The relative intensities of the signals corresponded to approximately 20% dimer-like and 80% substrate-bound giving rise to a Kd_{obs} for the TF interaction with α-synuclein(Luc 87-100) of 170 μM. No difference in affected residues is detected between the H_6-TF and TF interaction with α-syn(Luc 87-100) indicating that the hexa-histidine tag on TF does not affect the nature of the interaction with α-synuclein (Fig 4.7.a).

4.4.4 TF dynamics in the presence of α-synuclein (Luc 87-100)

The study described in Section 3.5.4 found that the TF dimer is a highly dynamic species. In this Section, the effect of the TF interaction with a substrate protein on the dynamics is investigated in order to described potential characteristic dynamic properties of the TF substrate-bound state that might provide insights into the mode of substrate interaction and chaperoning mechanism.

4.4.4.1 EPR

Electron Paramagnetic Resonance (EPR) spectroscopy was initially pursued as a means of obtaining long range distance constraints in the TF dimer. A theoretical introduction to DEER measurements by EPR and experimental details are provided in Section 2.5.1. This study was undertaken in collaboration with Dr Enrico Salvadori and Prof Chris Kay.

As was shown in Section 3.5.4 however, the TF dimer displays significant dynamic behaviour. No clear distances between nitroxide labels could be detected using standard double electron-electron resonance (DEER) experiments of 300 μM proxyl labelled TF in deuterated Tico at pH 7.5 with 50% glycerol as shown in Fig 4.8.

Electron spin coupling was detected as shown by the rapid decrease in echo intensity at early time points (Fig 4.8.a and b) indicating that the two nitroxide spin labels at position 326 in each TF monomer within the dimer were in proximity of each other over the course of the experiment. Efficient TF dimerisation was thus maintained upon freezing of the

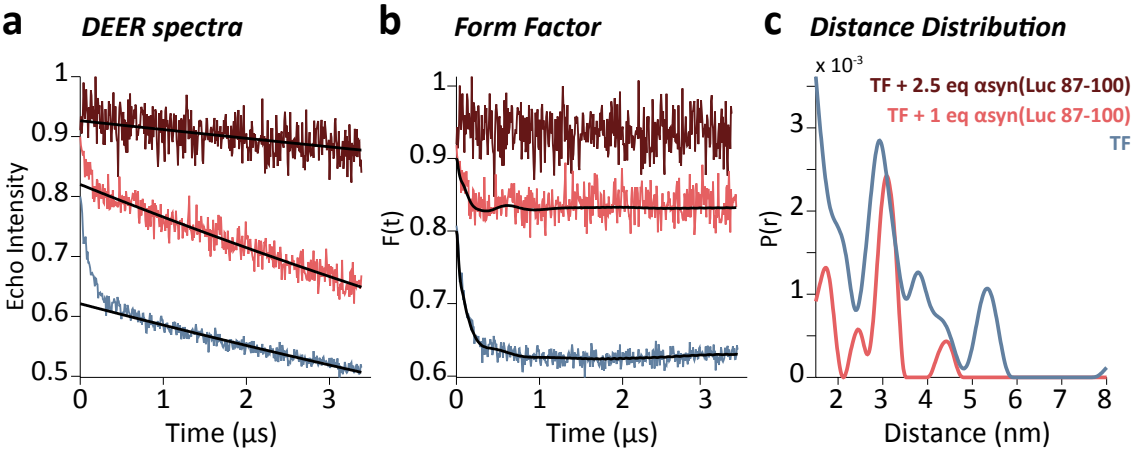

Figure 4.8: EPR study of the TF interaction with H_6-αsyn (Luc 87-100). (**a**) Double electron-electron resonance (DEER) traces of 300 μM TF labelled at position 326 in the absence (blue), at 1 equivalence (pink) and 2.5 equivalence α-synuclein(Luc 87-100). Estimated background relaxation is represented by a solid black line (**b**) DEER traces after subtraction of background relaxation. Fits of the experimental data are shown in solid black lines. (**c**) Spin distance distribution calculated from the fits of the experimental data in b.

sample. The absence of clear discernible oscillation in the signal suggests that multiple distances between the two spin labels are detected as shown in Fig 4.8.c. The calculated distance distribution covers a large range from short 1-2 nm distances up to 5.5 nm. The accuracy of the distance distribution presented is low as the sinusoidal behaviour of each of the detected distances interfere. The DEER experiments of TF 326C thus corroborate the structural heterogeneity of the TF dimer.

Gradual addition over two concentration points of H_6-α-syn(Luc 87-100) to spin labelled TF 326C resulted in the disappearance of the detected spin coupling as shown in Fig 4.8.b. At equimolar concentrations of TF and α-syn(Luc 87-100) ~33% residual electron spin coupling could be observed with no significant changes to the calculated distance distribution. The specific contribution from spin coupling to the DEER signal however appears absent at 2.5 fold excess of α-syn(Luc 87-100) suggesting saturation of TF is achieved. The gradual disappearance of the detected spin coupling indicates significant distancing of the two spin labels upon interaction with the H_6-α-syn(Luc 87-100) substrate suggesting the interaction with the substrate results in monomerisation of TF.

4.4.4.2 Carr-Purcell Meiboom-Gill relaxation dispersion

CPMG relaxation dispersion measurements of dimeric TF were described in Section 3.5.3.2. Here, they were undertaken at 800 MHz, 25°C and pH 7.5 in D_2O, with 2.5 fold excess α-syn (Luc 87-100) added to 400 μM TF^{ILV} in order to investigate the millisecond dynamics of the substrate-bound state. Dispersions were detected for 12 out of 66 resonances

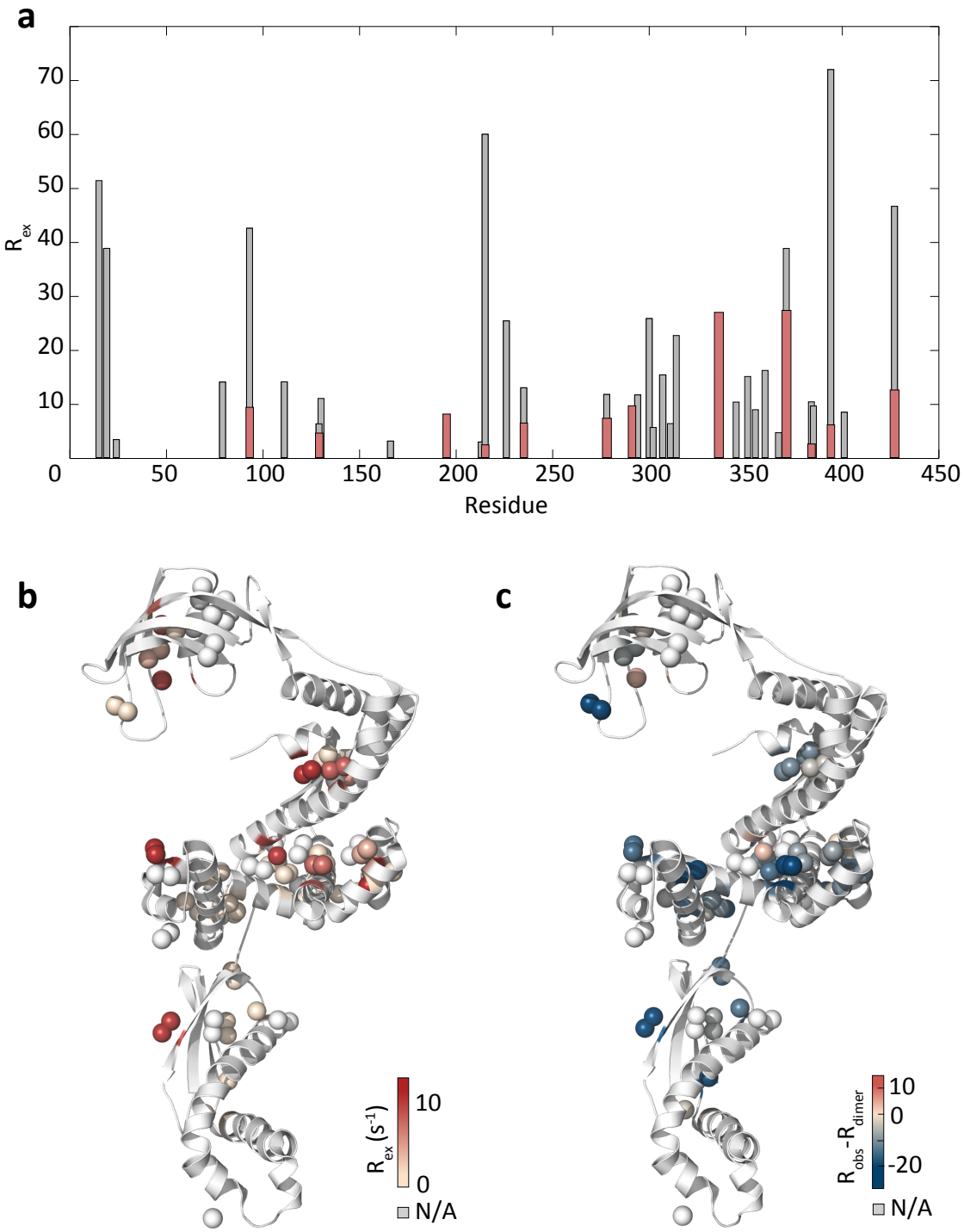

Figure 4.9: TF^{ILV} CPMG relaxation dispersion in the presence of H$_6$-α-syn (Luc 87-100). (**a**) Exchange contribution to the transverse relaxation (R_{ex}) of TF^{ILV} resonances in the absence (grey) and presence of excess H$_6$-α-syn (Luc 87-100) (pink) as determined by CPMG relaxation dispersion measurements at 25°C and 800MHz. (**b**) Structural distribution of R_{ex} values in the presence of H$_6$-α-syn (Luc 87-100) shown in a. (**c**) Difference in R_{ex} values in the absence and presence of H$_6$-α-syn (Luc 87-100).

monitored although the measured R_{ex} values were small for the majority (<10 Hz) with notable exceptions for residues 335 and 375. The identified residues displaying dispersion in the presence of substrate form a subset (33%) of those identified in dimeric TF while no new dispersions were detected. A comparison of R_{ex} values in the TF dimer and substrate bound state is presented in Fig 4.9 along with the spatial distribution of detected dispersions.

Generalising, the addition of the α-syn (Luc 87-100) substrate resulted in the disappearance or reduction of exchange dispersion R_{ex} values. The simplest explanation for the observed changes is the monomerisation of TF upon interaction with H_6-αsyn (Luc 87-100) effectively disrupting the dynamics of dimer formation described in Section 3.5.3.2. The observed dispersions would then originate from a residual population of TF dimer indicating that under the current experimental conditions (400 μM TF and 1 mM H_6-α-syn (Luc 87-100)), the binding event is not fully saturated. The value of R_{ex} is dependent on a large number of variables such as the population distribution, the exchange rate and the chemical shift difference between states (see Section 2.1.6.2). Therefore, they do not scale linearly with changes to the population distribution. The absence of an apparent trend to the reduction in R_{ex} values detected in the presence of substrate therefore is not in contradiction with the reduction of the dimer TF population upon interaction with H_6-α-syn (Luc 87-100). The possibility that the detected dispersion reflect substrate-interaction specific millisecond dynamics can however not be conclusively dismissed.

4.4.5 Discussion

The sensitivity of NMR spectroscopy to transient interactions in particular when associated with dramatic changes in apparent size has permitted the identification of a pre-existing TF interaction site in the native α-synuclein sequence of weak affinity centred around residue Phe 4. This N-terminal interaction is enhanced by the presence of a H_6-tag. Reduced temperatures appear to further stabilise this interaction suggesting the main driving force of the interaction is of non-hydrophobic nature. Although the detected interaction at the α-synuclein N-terminus is of low affinity (estimated at 4 μM from chemical shift perturbations), it may play an important role on the ribosome due to the high local protein concentrations.

The substitution of 14 α-synuclein residues with a sequence from firefly luciferase introduces a higher affinity interaction (estimated at 1 μM). This additional interaction does not appear to interfere with the pre-existing N-terminal binding even at high protein concentrations and additionally intensifies secondary interaction sites along the sequence. It is thus likely that H_6-αsyn (Luc 87-100) interacts with multiple TF sites. The α-synuclein regions involved in interaction with TF are enriched in aromatics with interactions cen-

tred around Phe, Tyr and His residues. Charged residues, particularly prevalent in the 40 C-terminal residues ofα-synuclein, appear on the contrary to interfere with the TF interaction.

Excess concentrations of H_6-αsyn (Luc 87-100) results in monomerisation of TF with an apparent K_d of 140 μM. The limited changes in the H_6-TF^{ILV} ^{1}H-^{13}C HMQC spectrum upon addition of WT H_6-αsynuclein suggests the N-terminal interaction alone is not sufficient to induce monomerisation. The results presented thus indicate that TF interacts with unfolded protein substrates in both dimeric and monomeric forms. Multiple TF binding sites in the substrate sequence and/or high affinity sequences appear required to induce monomerisation of TF at protein concentrations well above the K_d of dimerisation (8 μM). Furthermore, the observed monomerisation in the presence of a relatively strong affinity substrate (H_6-αsyn (Luc 87-100)) suggests competition exist between substrate interaction and dimerisation.

4.5 TF interaction with inactive ribosomes

The interaction of TF with protein substrates on the ribosome is mediated by binding to L23 on the 70S [126]. The interaction with L23 is necessary and sufficient for recruitment to the ribosome surface [137]. The TF interaction with the 70S induces a conformational change, most notably in the RBD but further elongation of the TF structure was detected by FRET [148]. The characterisation of isolated protein interactions thus cannot provide a full understanding the TF interaction with nascent chains. The study of the structural and dynamic behaviour of TF on the 70S particle in solution would provide such initial insights.

With reported affinities ranging from 100 nM - 1 μM and residency times of 100 ms to 10 s depending on the experimental methods [148, 162], the interaction of TF with the 70S ribosome would be predicted to be in the intermediate to slow NMR exchange regime. The exact exchange regime, further influenced by the chemical shift difference between the free monomer and the 70S-bound states will in part determine the likelihood of visibility of the ribosome-bound state by NMR (Section 2.1.6.1)

The flexibility of TF on the ribosome will further influence the observability of TF bound to the ribosome. Since NMR signals from the ribosome uniquely map to flexible regions such as L7/L12 and S1 [183, 184], TF would need to maintain significant flexibility on the ribosome to be detected by NMR, even in a favourable exchange regime such as slow exchange. Various crystal structures (Fig 3.22, [84, 126, 238]) indicate that inter-domain flexibility exists and conformational freedom of 10° in all directions around the TF ribosome binding motif and ribosomal protein L23 interaction has been reported [126]. The following study explores the possibility of studying TF on the ribosome and its interaction

with nascent chains.

4.5.1 TF binding to ribosomes result in general broadening of the NMR signal

Chemical shift perturbations: Addition of equimolar 70S ribosomes to 10 μM H_6-TF^{ile} in deuterated Tico buffer, pH 7.5 at 25 °C resulted in significant changes to the ^{1}H-^{13}C HMQC resonances as shown in Fig 4.10.a. Splitting of a number of resonances is observed, suggesting exchange between multiple TF states and a number of novel cross-peaks of uncertain origin are detected. While most resonances can be attributed to either monomeric or dimeric TF by comparison to the deconvoluted spectra presented in Section 3.5.2.3, two intense resonances at 0.82 and 12.5 ppm and 0.84 and 12.7 ppm in the ^{1}H and ^{13}C dimensions respectively cannot be identified.

Resonance intensity analysis : H_6-TF^{ile} ^{1}H-^{13}C HMQC resonances are broadened selectively with the strongest signal reduction to 40% of the intensity in the absence of the ribosome. According to the mathematical model developed in Section 3.4, at 10 μM TF to 10 μM 70S, 66 % of the total TF concentration is predicted to be bound to the ribosome. The interaction with ribosomes results in a shift in the monomer/dimer equilibrium. The population of the two states is shifted from 54 % dimeric and 46 % monomeric to 12 % dimer and 22 % monomer respectively while the remainder is ribosome-associated. Based on the calculated change in the population distribution, an initial prediction of the total combined monomer and dimer contribution to the NMR signal in the presence of the ribosome corresponds to 38 % at 10 μM H_6-TF^{ile} in close agreement with the strongest broadening events in the presence of equimolar ribosome.

The prediction of the combined monomer and dimer contribution is based on the assumption that the TF monomer contributes twice the signal intensity due to its smaller size further discussed in Section 3.5.1.1. In the absence of ribosomes, 5.4 μM dimeric TF contributes once and the 4.6 μM monomeric fraction contributes twice, resulting in a total signal intensity of 146% the signal of 10 μM pure TF dimer. In the presence of ribosomes, the shift in the dimerisation equilibrium produces an expected 56% of the 10μM pure dimer intensity. The ratio of these two predicted intensities gives rise to an expected signal reduction to 38%.

Prediction of TF resonance intensities The deconvolution of the TF monomer and dimer spectrum undertaken in Section 3.5.2.3 permits the prediction of the appearance of the TF ^{1}H-^{13}C HMQC spectrum in the presence of the ribosome, with the assumption that the ribosome-bound form of TF does not contribute to the detected signal through the absence of motional freedom upon interaction with the 70S particle. The predicted

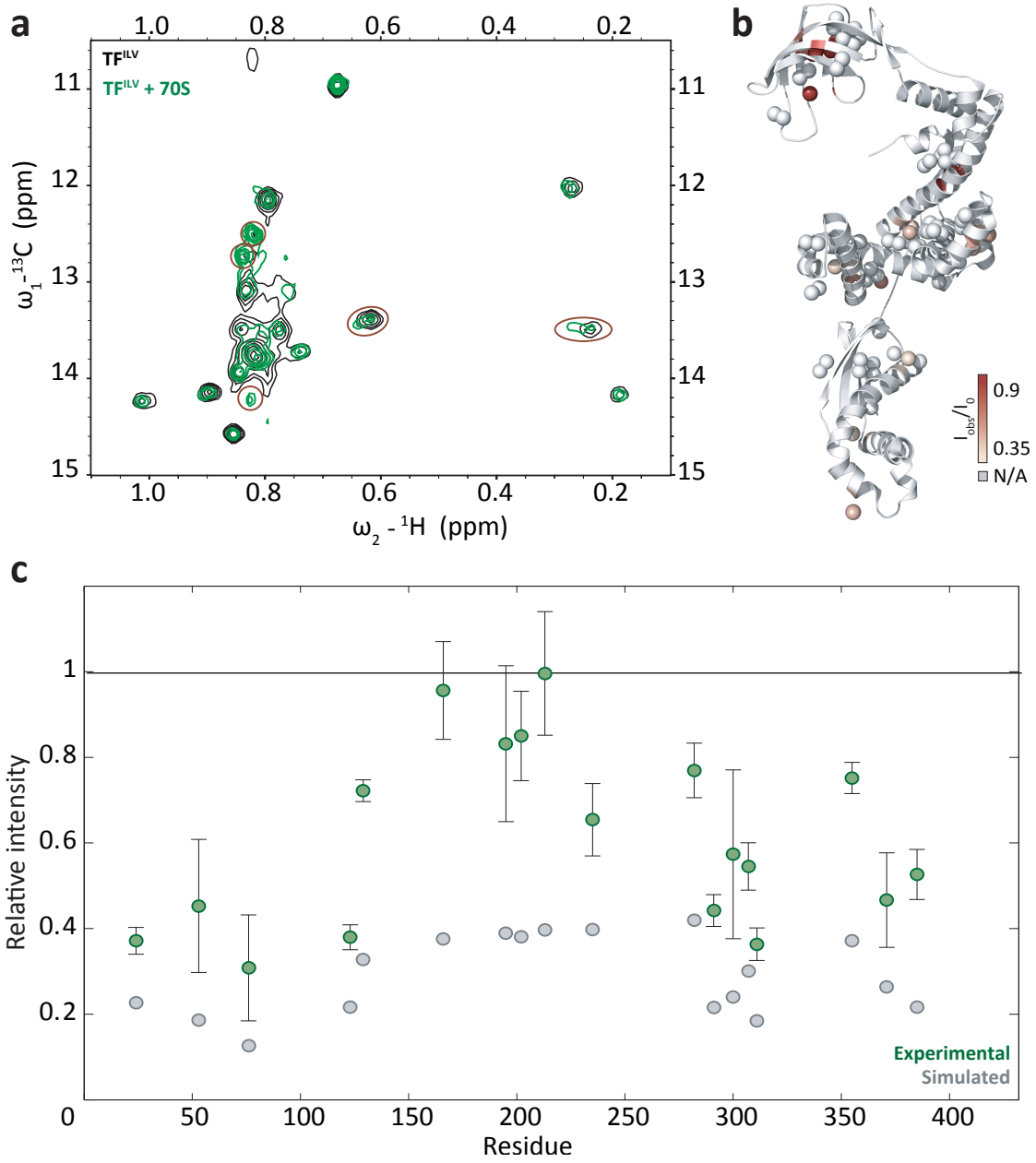

Figure 4.10: TF interaction with 70S ribosomes (**a**) ^{1}H-^{13}C HMQC spectrum of 10μM H_6TF^{ile} in the absence (black) and presence of 10 μM 70S ribosomes (green). Resonances that display significant perturbations are circled in red. **b**) Structural distribution of relative resonance intensities in the presence of ribosomes as shown in c. (**c**) Resonance intensities in the presence of equimolar ribosome relative to 10μM H_6-TF^{ile} only (green). Predicted relative signal intensities in the absence of contribution from the ribosome-associated TF state, based on the deconvoluted monomer and dimer spectra, Section 3.5.2.3 (grey).

signal intensity reductions based on the dilution study of TF^{ILV} at 900MHz are shown in Fig 4.10.c. The detected NMR intensity for specific ^{1}H-^{13}C HMQC resonances is up to 2.5 fold stronger than predicted values.

The structural distribution of the observed resonance intensities is shown in Fig 4.10.b. Resonances least affected by broadening (80-100% signal intensity) in the presence of ribosomes map to the TF PPIase domain. The central SBD further shows resonance intensities (40-70%) above and beyond expected signal intensities in the absence of the contribution of the 70S-bound state to the NMR signal.

4.5.2 Discussion

The addition of equimolar ribosomes to TF^{ile} results in changes to both the position of ^{1}H-^{13}C HMQC resonances and their measured intensities. While the changes in the dimerisation equilibrium make the analysis of these perturbations not straightforward, the mathematical model developed in Section 3.4 and the deconvoluted spectra of 1 μM dimeric and 1 μM monomeric TF determined in Section 3.5.2.3, allow the prediction of resonance intensities in the absence of contribution from the ribosome-bound state of TF. The disagreement between the predictions and the experimental data suggest that TF while bound to the ribosomal surface does in fact contribute to the observed spectrum.

The current intensity data suggests the TF domains furthest removed form the ribosome surface, in particular the PPIase domain, may be observable by NMR when bound to the ribosome. The extent of broadening observed in the RBD closely matches the predicted intensity values. While the signals from this domain when bound at the ribosome surface are thus broadened beyond detection, the absence of additional exchange broadening indicates that the TF interaction with the 70S ribosome is in the slow exchange regime, creating favourable conditions for the study of TF at the ribosome surface.

The differences in TF construct and magnetic field strength between the TF/70S experiment described here (H_6-TF^{ile}, at 700 MHz) and the dilution experiment described in Section 3.5.2 (TF^{ILV} at 900 MHz) mean that the observability of TF cannot be conclusively determined. Uncertain is to what extent the H_6-purification tag interferes with the dimerisation equilibrium which could introduce errors in the prediction of the various TF populations present here. The magnetic field strength influences the regime of the exchange event such that a reaction in slow exchange at high field might present intermediate exchange at lower field strengths. Further experiments will need to be undertaken to determine the feasibility of studying TF directly on the ribosome by NMR.

In particular, deuteration of the 70S together with a reduction in the TF concentration relative to the ribosome could improve the quality of the TF ^{1}H-^{13}C HMQC spectrum by populating the bound-state to a greater extent and by reducing the relaxation associated

with domains in close proximity of the ribosome surface. Matching the experimental conditions of the TF dilution outlined in Section 3.5.1.1, will permit a more ready interpretation of the obtained data.

4.6 TF interaction with α-synuclein ribosome nascent chain complexes

The following preliminary investigation of the interaction of TF with the α-synuclein RNC considers the behaviour of the nascent chain. This investigation was undertaken in close collaboration with Annika Weise (UCL). Previous studies of the TF interaction with α-synuclein RNC found that while the nascent chain is in proximity of TF, it was not protected from proteolysis by the presence of TF at the ribosomal exit tunnel [159]. These results indicated that the α-synuclein nascent chains maintain significant flexibility on the ribosome. The α-synuclein RNC therefore might remain observable by NMR in the presence of TF, enabling a residue specific characterisation of the effect of TF on the dynamics of the nascent chain.

4.6.1 Summary of the α-synuclein RNC NMR intensity data

^{15}N labelled α-synuclein RNC (α-syn RNC), prepared by Annika Weise (UCL), gives rise to a SOFAST HSQC spectrum of remarkable quality (Weise et al, in preparation). The spectrum shows limited proton dispersion, characteristic of an unfolded protein, indicating α-synuclein does not adopt a folded conformation on the ribosome. Broadening is observed for the whole nascent chain down to 10% intensity with signals of the N-terminal region of the sequence attenuated beyond detection. This observation suggests the α-synuclein nascent chain interacts transiently with the highly charged ribosome surface. Addition of TF^{ILV} to the α-syn RNC results in further broadening of the first 100 residues to approximately 25% of their intensity in the absence of TF (Weise et al, in preparation, Fig 4.11). Although no specific interaction of TF with WT α-syn RNC has been described, the presence of TF at the ribosomal exit tunnel does appear to affect the conformational freedom of the α-synuclein RNC. A gradual titration of TF to the α-synuclein RNC was pursued in order to characterise the TF/RNC interaction equilibrium in further detail. The quantitative analysis of the acquired titration data and a description of the associated difficulties in the interpretation of this data are described below.

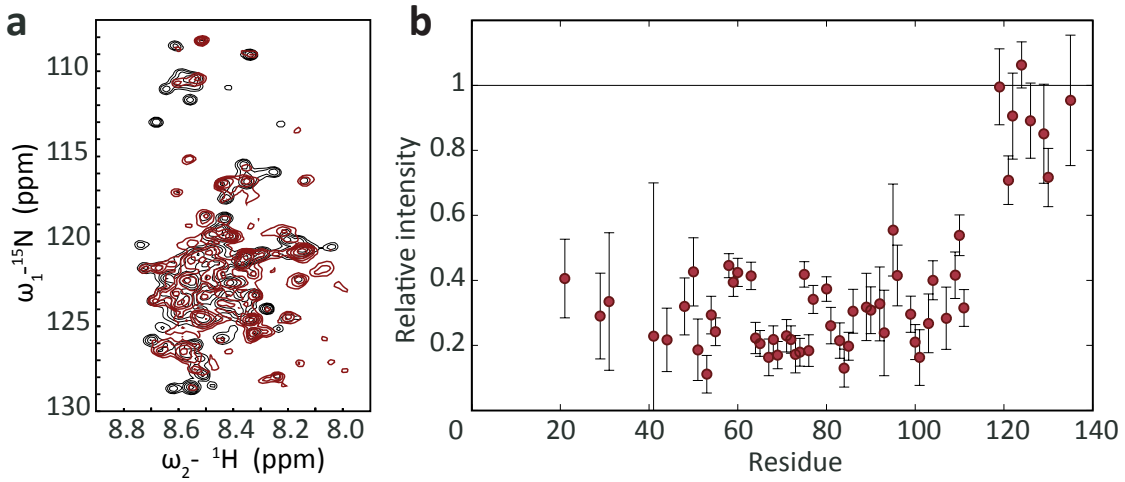

Figure 4.11: Interaction of TF with the α-synuclein RNC. (**a**) ^{15}N SOFAST HMQC spectrum of the α-syn RNC in the absence (black) and presence of one equivalence TF (red). (**b**) Relative ^{15}N signal intensities of α-syn RNC resonances in the presence of one equivalence TF.

4.6.2 Quantification of ^{15}N titration data

The study of the TF interaction with the α-syn RNC was undertaken using 6.5 μM α-synuclein RNC, occupied to 80% as determined by immuno western blot assay, 5 μM 70S and TF^{ILV} concentrations ranging from 1 to 25 μM at pH 7 and 4°C. The signal of the first increment of the ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N labelled α-synuclein RNC was integrated and the amplitude of the signal monitored during gradual addition of TF over 7 points ranging from 0 to 4 fold molar equivalence (Fig 4.12). Integrated signals were corrected for dilution of the sample over the course of the titration using the reference signal of DSS at 0ppm in ^{1}H 1D NMR spectra recorded at each point of the TF titration. 0.01% DSS was added at the start of the titration only, such that the decrease in its signal intensity reports directly on the dilution of the sample. Errors in the signals were estimated by integration of noise in an identically sized spectral region.

The ^{15}N signal of 70S ribosomes in the absence of nascent chains is shown in Fig 4.12.a and b. It is unaffected by the addition of TF showing clearly that the observed changes in the ^{15}N signal of the α-synuclein RNC reflect changes in the nascent chain and do not correspond to transient interaction of TF with other flexible (NMR observable) regions of the ribosome such as the L7/L12 stalk region.

The ^{15}N signal integrals (I_{obs}) were normalised to the initial ^{15}N (I_{free}) signal in the absence of TF for ease of visualisation and comparison with the equivalent 70S experimental data. The measured intensities can be described as the sum of the intensity of the non-TF bound fraction of RNC (I_{free}) and the corresponding intensity of TF-bound RNC

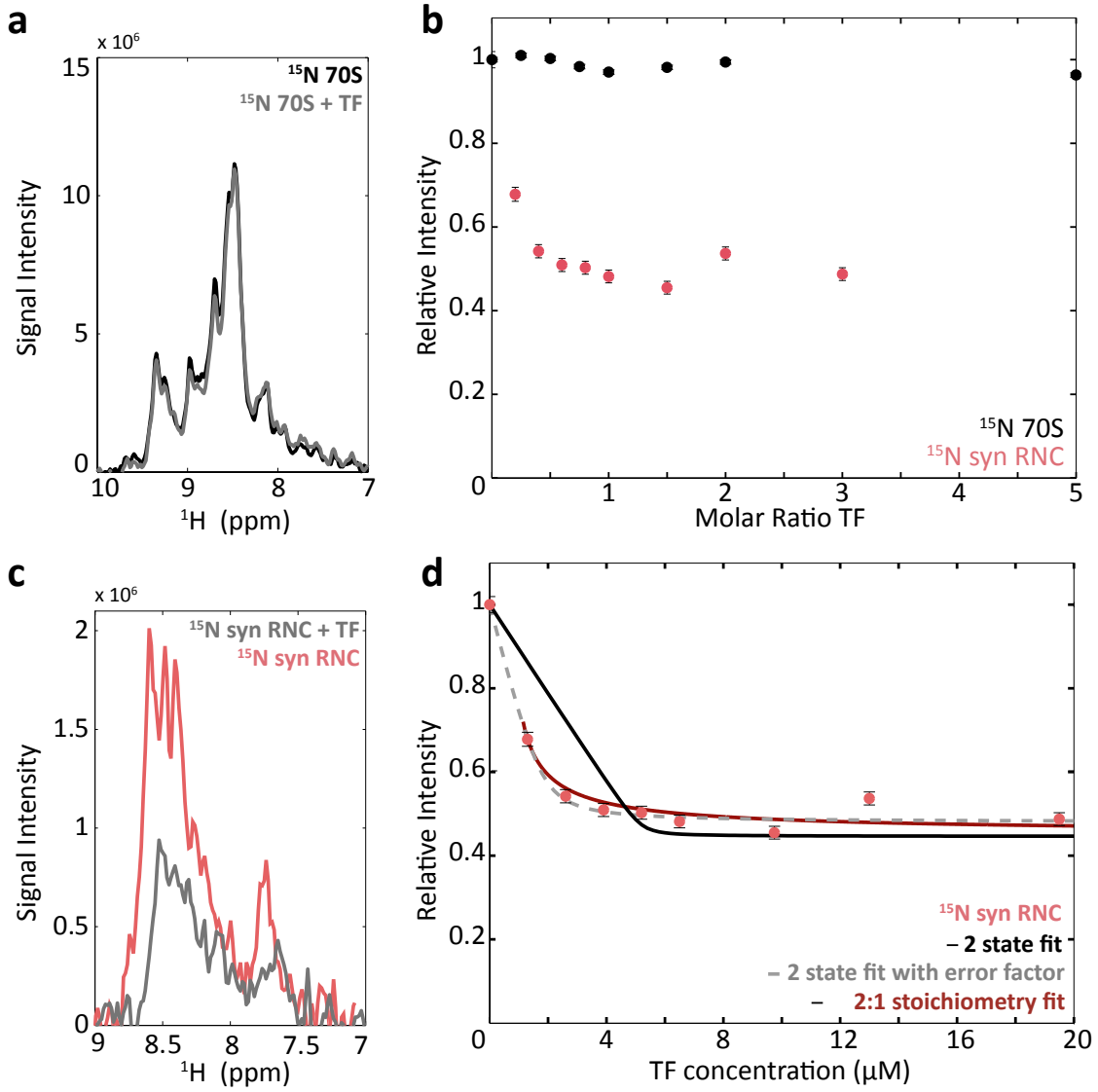

Figure 4.12: Quantification of ^{15}N TF titration data. (**a**) Amide region of first increment of ^{15}N SOFAST HMQC spectrum of 70S ribosome in the absence (black) and presence of 1 equivalence TF^{ILV} (grey). (**b**) Relative ^{15}N signal integral intensities 5 μM 70S (black) and 6.5 μM α-synuclein RNC (pink) with increasing concentrations of TF^{ILV}. (**c**) Amide region of first increment of ^{15}N SOFAST HMQC spectrum of 6.5μM α-synuclein RNC in the absence (pink) and presence of 1 equivalence TF^{ILV} (grey).(**d**) Numerical solution to two state equilibrium fitted to the experimental signal intensity ratios without (pink), with factor accounting for experimental errors (grey dotted) and with 2:1 stoichiometry (red).

(I_{bound}) where P_{free} and P_{bound} describe the TF population distribution at each stage of the titration:

$$I_{obs} = I_{free} \times P_{free} + I_{bound} \times P_{bound} \tag{4.6.1}$$

which after normalisation gives:

$$\frac{I_{obs}}{I_{free}} = 1 \times P_{free} + \frac{I_{bound}}{I_{free}} \times P_{bound} \tag{4.6.2}$$

where

$$P_{free} + P_{bound} = 1$$

The population distribution (P_{free} and P_{bound}) was then obtained by solving the quadratic equation describing the two state TF/RNC interaction equilibrium (see Section 3.4) for the TF-bound RNC population (P_{bound}=[TF:RNC]/[R_0]):

$$\begin{aligned} Kd &= \frac{[TF] \times [RNC]}{[TF:RNC]} \\ T_0 &= [TF] + [TF:RNC] \\ R_0 &= [RNC] + [TF:RNC] \end{aligned} \tag{4.6.3}$$

The solution to the above simultaneous equations was fit to the observed intensity data. The obtained best fit is shown in Fig 4.12.d (solid pink line). Initially, just two parameters were used for the fit, the dissociation constant (K_d) and the normalised NMR intensity of the TF-bound RNC ($I_{obs}/_{free}$). It is apparent that the data cannot be accurately described by the fitted equation as shown in Fig 4.12.d.

The solution to the least squares minimisation assumes that all experimental factors such as protein and ribosome concentrations and ribosome occupancy are accurate. Introduction of an 'error' factor on these experimental constants significantly improves the quality of the fit (Fig 4.12) but suggests an error of at least 30% should be taken into account for each experimental factor. Experimental errors of this amplitude can readily be introduced in the study of the ribosome where low particle concentrations give rise to large absorbance values and multiple dilution steps are required for concentration determination. Moreover, ribosome occupancy is difficult to determine due to the technical limitations of densitometry analysis. The inclusion of this error factor visually improves the quality of the fit yet the determined dissociation constant values remain associated

with errors of 100% or more. The additional parameter introduced in the form of an error factor leaves the fitted equation under-restrained.

The fit of the presented data to a 2:1 TF to RNC stoichiometry (Fig 4.12.d) significantly improves the results, estimating the value of the dissociation constant of the reaction at 1.15 ± 0.08 μM. There is however no experimental evidence to justify the proposed binding model. The improved fit is consistent with a systematic error in the concentration of one or both the reactants. An accurate quantification of the dissociation constant could be achieved, in contrast to the inclusion of an error factor which similarly reflects inaccuracies in the stock concentrations, because the number of fitting parameters is limited to two.

Discussion The titration of ^{15}N α-syn RNC demonstrated that the gradual addition of TF leads to the progressive broadening of the NMR signal while the NMR signals associated with the 70S are unaffected, indicating that the residue specific broadening reported reflects the interaction of TF with the nascent chain only. The monitoring of 1D signal integrals over the course of a titration allows the rapid acquisition of a titration curve where sample stability is limited. Quantitative analysis of such titration data is possible following the strategy outlined above. In the case presented here, the introduction of an experimental error factor was required to produce satisfying fit results since the detected broadening saturated below stoichiometric concentrations. While the amplitude of the errors determined are feasible, it is possible that the binding model chosen does not accurately reflect the monitored reaction.

4.6.3 Quantification of ^{13}C titration data

Over the course of the titrations described above, the added TF^{ILV} was monitored by interleaving single increment ^{1}H-^{13}C HMQC spectra between the ^{15}N SOFAST HSQC spectra of α-syn RNC described. Representative examples of the spectra obtained are shown in Fig 4.13.a and b. The signal in these spectra was integrated over the methyl region (0-1.2 ppm) and using a similar approach to that used for the ^{15}N RNC data, were monitored at each point of the TF titration. Signals were corrected for dilution by normalisation to the DSS peak at 0 ppm. Errors were estimated by integration of an identically sized distant region of the spectrum. Since ^{13}C is a naturally occurring carbon isotope, small amounts of background labelling of ribosomes and nascent chain were detected. ^{13}C integral intensities was normalised by background subtraction using the intensity of the ^{13}C signal in the absence of TF.

A key assumption that was made in the interpretation of this data is that the ribosome-bound TF state is unobservable by NMR. This assumption is based on the idea that the size of the ribosome bound form would be too large for detection by NMR even while exploiting the advantages of methyl-TROSY NMR. The results from Section 4.5 contradict this

assumption to some extent by showing significantly more signal intensity in the presence of ribosomes than expected based on the re-arrangement of the monomer/dimer equilibrium only. The implications of this assumption will be addressed in the discussion 4.6.3.3.

4.6.3.1 70S ribosome interaction with TF

The ^{13}C signal integrals for increasing concentrations of TF are shown in Fig 4.13.c. The titration data of TF^{ILV} with 70S were initially fit to the equation of a two-state equilibrium, which describes the interaction with the ribosome as the interconversion of two states, free TF and bound TF. The detected signal was assumed to originate from non ribosome-bound TF only in Fig 4.13.d (left panel). This yielded a K_{d} of 1.65 $\pm$ 0.7 μM for the TF/70S interaction in agreement with previously reported affinities summarised in Table 3.1, Section 3.4.1. It should however be noted that the ^{13}C titration data discussed in what follows was recorded alongside the ^{15}N signals discussed in Section 4.6.2. There, we found a large systematic error, most likely introduced through inaccuracies in the determination of stock protein concentrations. The results discussed below will be similarly affected by this systematic error and the values of the dissociation constants mentioned are likely to present under-estimates of the actual value.

Our current understanding of the TF equilibria indicates that the monitored reaction is not a simple two state equilibrium as the TF dimerisation reaction directly competes with the ribosome interaction. The 'free' TF signal detected thus has two different origins, monomeric and dimeric TF each with their specific contribution to the NMR signal as established in Section 3.4.3. The TF monomer was found to give rise to significantly more intense signals than the TF dimer ($\sim$ 2 fold) with a dimerisation K_{d} of 8 μM at 25°C under similar experimental conditions to those used here. Fitting of the three state equilibrium to the titration data using the experimentally derived constraints on the relative monomer to dimer signal (2 fold) and the value of the dissociation constant for dimerisation (8μM) produced a K_{d} value of 0.32 $\pm$ 0.30 μM for the TF interaction with the ribosome. Accounting for the competition of the dimerisation reaction thus results in a tightening of the calculated affinity. It should be noted than the error on this result is significantly larger that for the two state model (relative error increased from 42% to 94%) suggesting the fit is under-restraint. Moreover, visual inspection of the fit (Fig 4.13.d, right panel) suggests that this fit is of poorer quality than that of the two state model; the shape of the fitted curve, imposed by the previously determined restraint on the monomer to dimer intensity does not agree with the experimental data. Release of this constraint on the monomer intensity visually improves the fit of the data to the proposed model (Fig 4.13.d, middle panel). The error in the K_{d} is however more than 1000% which reflects the fact that, again, the equation is under-restraint because an additional parameter has been

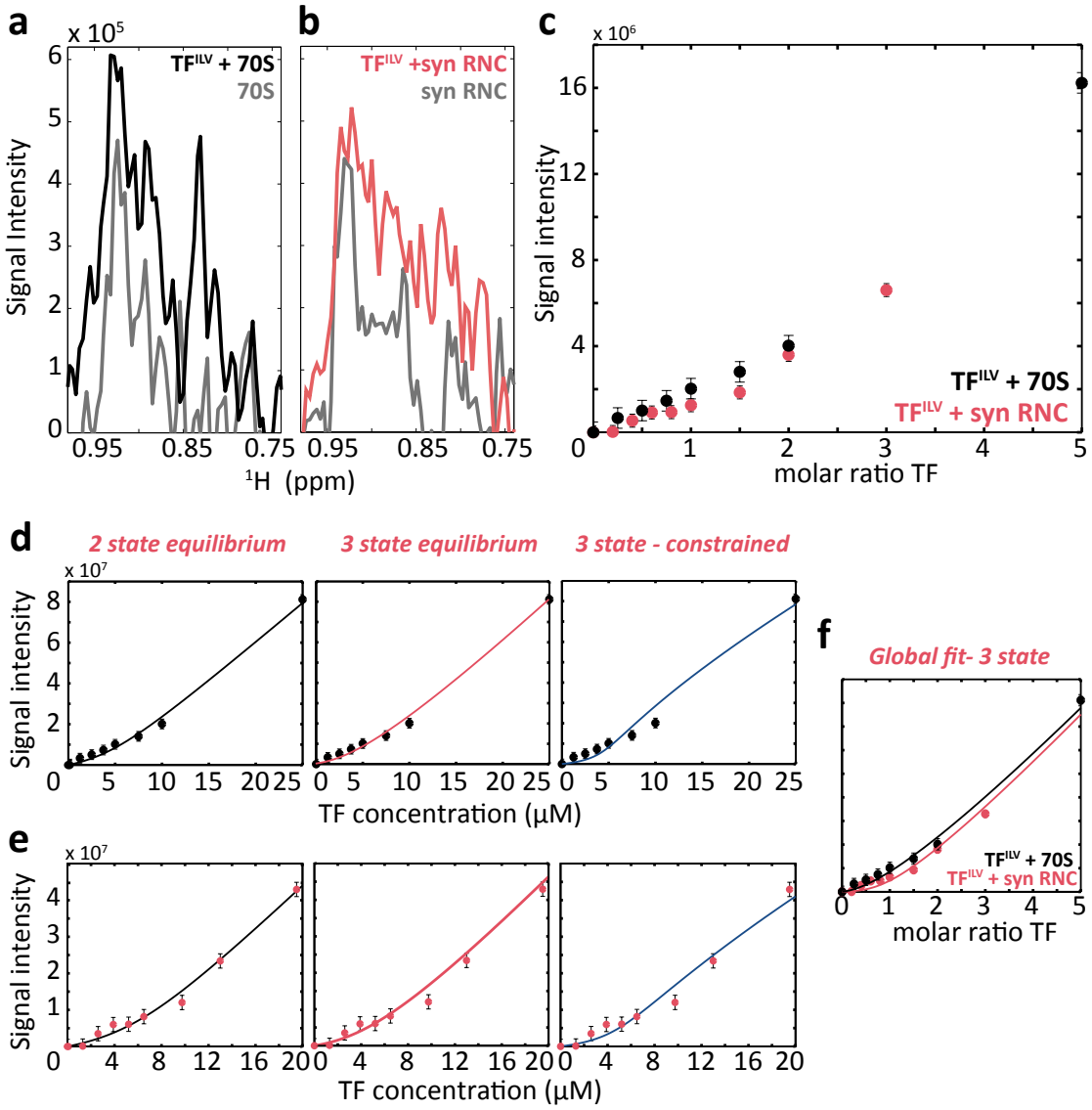

Figure 4.13: Quantification of ^{13}C TF titration data with the 70S ribosome and the α-synuclein RNC. (**a**) Methyl region of the first increment of ^{13}C HMQC spectrum of 5μM 70S ribosome in the absence (black) and presence of 1 equivalence TF^{ILV} (grey) (**b**) Methyl region of first increment of ^{13}C HMQC spectrum of 6.5μM α-syn RNC in the absence (pink) and presence of 1 equivalence TF^{ILV} (grey) (**c**) Relative ^{13}C signal integral intensities of TF^{ILV} in the presence of 5 μM 70S (black) and 6.5 μM α-synuclein RNC (pink). Errors were estimated by integration of a same sized region of noise in the NMR spectrum. (**d**) Individual fits of the 70S and (**d**) α-synuclein RNC ^{13}C intensity data to the mathematical solution of a two-state (black), three-state (pink) and constrained three-state (blue) equilibrium as described in the text. (**f**) Fits to the experimental signal intensities based on a global analysis approach described in the text.

introduced in the form of a factor describing the relative intensities between monomeric and dimeric TF. The intensity ratio of monomer to dimer fit was 0.15. What becomes evident is that the TF monomer fraction contributes significantly less to the observed signal in the presence of ribosomes as would have been expected from the study of the TF dimerisation reaction in isolation where the TF monomer presented twice the intensity of the dimer.

One possible explanation for the remarkable reduction in the observed monomer signal contribution is the transient, non specific, interaction of the TF monomer with the ribosome surface under the experimental conditions used here. This idea is corroborated by the observation that the detected TF signal at 5:1 equivalence TF to 70S (25 μM TF to 5 μM 70S) is significantly less intense than that observed for 10 μM TF in isolation. TF interacts specifically at the ribosomal exit tunnel in a 1:1 stoichiometry such that at 5:1equivalence 20 μM of TF is expected to be free in solution. In NMR terms, the latter observation could also be attributed to quirks in the exchange timescales of the TF- ribosome interaction resulting in intermediate exchange broadening of the signal. With a reported half-life of 15 sec for the TF/70S complex the NMR exchange timescale is expected to be in the slow exchange regime. However, the recent report of a half-life of 100 ms only [162] could support an intermediate exchange regime.

In summary, the ^{13}C titration data can be successfully fitted to a two-state binding equilibrium resulting in a $K_{\mathrm{d}}^{\mathrm{obs}}$ of 1.65 $\pm$ 0.7 μM in agreement with previously determined values for this two-state equilibrium. A more accurate description of the monitored reaction accounts for the competing dimerisation reaction where the TF monomer and dimer contribute independently to the detected NMR signal. In the absence of the ribosome, the TF monomer was shown to contribute twice the signal of same concentration of the TF dimer due to the differences in molecular size. The inclusion of this information and the experimentally derived dissociation constant for dimerisation did not introduce additional parameters to the equation but failed to described the experimental data accurately. The unrestraint fit of the monomer and dimer contribution individually, visually improved the quality of the fits but large errors indicated that the fit was under-restrained. The relative intensity of the monomer to dimer contribution of 0.15 is significantly lower than in the absence of the ribosome. This difference suggests that either the assumption of slow exchange does not hold or that an additional non specific interaction of the TF monomer with the ribosome surface was detected.

4.6.3.2 α-synuclein RNC interaction with TF

Following a similar approach as for the 70S, ^{13}C intensity data was gathered during the titration of the α-synuclein RNC with TF^{ILV}. As described above Section 4.6.3.2, signal

integral values were corrected for dilution and the background ^{13}C signal subtracted. The titration data was fitted to a binding model accounting for three TF states: free TF, ribosome-bound and RNC-bound TF states. The K_d for interaction with the 70S ribosome determined in Section 4.6.3.2 (1.65 μM) and a ribosome occupancy of 80% were used as input parameters to restrain the fit. The K_d obtained for the α-synuclein RNC was 1.69 $\pm$ 0.77 μM, within error of the experimentally derived K_d for the interaction with empty ribosomes, suggesting the presence of the α-synuclein nascent chain does not increase the affinity of TF for the ribosome.

Indeed, a plot of the 70S and RNC ^{13}C titration intensity data (Fig 4.13.c) reveals that both traces are within error of each other. However, the data for the α-synuclein appears to be consistently below the 70S data points hinting a potentially tighter interaction of TF with the α-synuclein RNC. As described for the 70S interaction with TF, the data were fit to a four state equilibrium accounting for the TF dimerisation reaction and differences in TF monomer and dimer signals (Fig 4.13). The same observation about the reduced contribution of the monomer to the NMR signal was made and attributed to the two possible sources described above: exchange broadening or a non specific interaction with the ribosome.

Global fit of ^{13}C data. Since the fit of more complex, more accurate, binding models appear under-restrained for both the TF interaction with the ribosome and α-synuclein RNC, a simultaneous fit of both the 70S and RNC ^{13}C data was undertaken. Global parameters include the intensities associated with the various TF states and the K_d values for the TF dimerisation and 70S binding reactions.

Once more, the simplest binding model was initially pursued. The 70S data was fitted to a two-state equilibrium (TF and 70S bound) and a three-state equilibrium (TF, 70S and RNC bound) for the ^{13}C RNC data. This least square minimisation approach resulted in K_d values of 2.4 μM and 0.53 μM respectively for the 70S and RNC interaction. Extension of the binding model to include competition with the TF dimerisation reaction produced dissociation constants of 6.7 μM and 1.1 μM respectively with a relative monomer to dimer intensity of 0.1 fold. The obtained fits to the experimental data are presented Fig 4.13.f. While the TF intensity differences between the 70S and the α-syn RNC appears small, this global fit approach consistently produces a 5 fold lower affinity of TF for the α-syn RNC over the ribosome.

4.6.4 Discussion

The quantitative analysis of the ^{15}N signal integrals of the 70S ribosome and α-synuclein RNC across various TF concentrations indicate the observed changes are specific to the TF-α-syn RNC interaction. However, the early saturation of the broadening (before one

equivalence) in resonance intensity suggests substantial experimental errors or incorrect assumption of the exchange regime of the reaction. As a consequence, the dissociation constant for the TF interaction with the α-synuclein RNC could not be determined.

The analysis of the ^{13}C signal intensities originating from increasing concentrations of TF^{ILV} added to the 70S ribosome found values for the apparent dissociation constant of TF interaction with the ribosome of 1.65 μM in agreement with previous studies. Attempts to account for the competition with the TF dimerisation reaction resulted in under-restrained fits. A detailed understanding of the TF dimerisation reaction was developed in Section 3.5.2.2. The application of this knowledge about the relative NMR intensity of the TF monomer and dimer and the dissociation constant of dimerisation showed that this information does not accurately described the mechanism of competition between dimerisation and ribosome interactions. It is likely that either a third non specific interaction of the TF monomer occurs at the ribosome surface or that the reaction is in the intermediate exchange regime suggesting relatively fast kinetics of TF interaction with the ribosome.

4.7 Concluding remarks

The sensitivity of NMR spectroscopy to transient interactions has permitted the identification of a pre-existing interaction between native α-synuclein and TF. This N-terminal interaction centred around Phe 4 is enhanced by the presence of a H_6-tag. The substitution of 14 α-synuclein residues with a sequence from firefly luciferase introduces a higher affinity interaction that does not appear to compete with the N-terminal interaction. The absence of apparent competition and the 'tethering' effect resulting in stronger broadening at Tyr 39 and His 50 in α-syn (Luc 87-100) relative to WT α-synuclein suggests multiple interactions occur simultaneously. It is thus likely that the various regions of H_6-α-syn(Luc 87-100) interact with individual TF sites. The interaction of TF with α-syn (Luc 87-100) is mediated through aromatic residues while charged residues in particular in the C-terminus of α-synuclein appear unconducive to the interaction. The behaviour of the various TF interaction sites on α-synuclein (Luc 87-100) across experimental conditions suggest differential properties of the multiple TF substrate-binding sites. The interaction of α-synuclein occurs with dimeric TF while the addition of α-synuclein (Luc 87-100) results in monomerisation. Unclear is whether the stronger affinity or the presence of multiple binding sites is at the origin of monomerisation. Substrate interaction with TF can thus compete with the TF dimerisation equilibrium.

H_6-TF^{ile} appears to maintain significant flexibility on the ribosome with the intensity analysis Section 4.5 indicating that regions distal from the ribosome surface in the TF-ribosome bound state such as the PPIase domain and to a lesser extent the SBD might

be observable by NMR. The extensive broadening observed in Section 4.6.3 for the TF^{ILV} interaction with 70S at 4°C however suggests a possible additional layer of complexity. At this point in time, it is difficult to attribute the differences in the behaviour of the TF signal across the two experiments to difference in the experimental set-up (temperature) or the nature of the TF construct used (H_6-TF^{ile} vs TF^{ILV}).

The analysis of residue specific signal intensities of the α-synuclein RNC in the presence of TF^{ILV} do not report on the presence of a specific N-terminal interaction as observed in the absence of the ribosome (Section 1.4.1) since the N-terminal resonances are unobservable in the RNC. The α-synuclein RNC is however significantly affected by the presence of TF as demonstrated by the reduction in the total signal intensity to 45%. Quantitative analysis of the 70S and α-synuclein RNC with TF^{ILV} is hindered by potential experimental errors, the complex TF equilibrium leading to an under-restrained system and an incomplete understanding of the observed phenomenon resulting in the over-simplification of the fitted model.

The current study presents the first detailed characterisation of the various competing TF equilibria in the presence of a single substrate, the α-synuclein RNC. It moreover highlights the difficulties associated with the interpretation of NMR data on such large and dynamic macromolecular systems and shows that further detailed characterisation of the individual reactions is required to fully understand the interplay between them.

Chapter 5

NMR investigation of the trigger factor interaction with ribosome-nascent chain substrates - ddFLN5

5.1 Introduction

In Chapter 4, the interaction of TF with the α-synuclein RNC was investigated as a null model for TF interactions with nascent chains at the ribosomal exit tunnel. The disordered nature of α-synuclein eliminates the complexity caused by partial folding of the substrate at the ribosome surface and has permitted an initial characterisation of the impact of TF on the available conformational space and the interaction of the nascent chain with the ribosomal surface.

In this Chapter, we complement this study by considering the interaction of TF with a folding competent protein model in its early stages of synthesis, with an aim to begin to formulate generalised statements about the nature of the interaction of TF at the exit tunnel. The protein model in question, ddFLN5, is a subdomain of a large multi-domain gelation factor (ABP120) and folds independently into an immunoglobulin-like β-sandwich (Fig 5.8.c, page 173).

The folding properties of ddFLN5 in isolation and on the ribosome have been extensively characterised in our group using NMR spectroscopy [65, 257]. In this regard, C-terminal truncation mutants of ddFLN5 have been used to mimic the progressive emergence of the ddFLN5 domain during translation (Waudby et al, in preparation). It was found that truncation of the final β-strand (G, Fig 5.8.c, page 173), corresponding to 16

truncated residues (ddFLN5Δ16) resulted in a disordered polypeptide across all sampled temperatures (10°C, 25°C and 37°C). The transition from this disordered state to the natively folded structure was shown to occur progressively with increasing protein length, via two native-like intermediate species. These folding intermediates appear in truncation mutants of Δ6, Δ8 and Δ9, with populations up to 50% relative to the disordered species for ddFLNΔ6 (Fig 5.1.a). In ddFLNΔ6, two folded conformations appear to be in slow exchange, as inferred from the observability of two sets of NMR resonances, indicating that *cis-trans* proline-isomerisation of P742 (9 a.a removed from the C-terminus) might be at the origin of the two folded states. This was supported by site-directed mutagenesis of P742 to alanine which biased the *trans* conformation of the peptide bond and demonstrated that the two intermediate populations are distinguished by rearrangements of proline isomerisation in the loop region preceding the final β-strand. In the ddFLN5Δ4 variant, with only 4 residues truncated from the full-length ddFLN5 protein, the final β-strand hydrogen bond network was observed to allow the full population of the native *cis* proline conformation. Indeed, its structural and dynamic behaviour, as studied by NMR, mirror those of the native state. The proposed, progressive, folding pathway of ddFLN5 is summarised in Fig 5.1.a .

On the ribosome, various nascent chain lengths of ddFLN5 have been studied with progressively longer segments of the subsequent domain 6 (ddFLN6) added to the ddFLN5 protein sequence (Cabrita *et al.*, *in preparation*, Cassaignau *in preparation*, 2014). At the shortest studied length (ddFLN5+21 RNC), the final β-strand, G, in the native ddFLN5 structure is predicted to be buried inside the ribosomal exit tunnel (estimated to be able to hold ∼30 amino acids in length, described Section 1.1.3.1) resulting in a folding incompetent state. The ^{1}H-^{15}N SOFAST HMQC spectrum of the ddFLN5+21RNC displays the limited ^{1}H chemical shift dispersion characteristic of an unfolded protein and closely resembles that of ddFLN5Δ16, the benchmark for the ddFLN5 unfolded state (Fig 5.2.a). Folding of the ddFLN5 RNC appears delayed, with the midpoint of folding around a linker length of 42-45 residues, where the full domain 5 sequence and an additional 12-15 residues are predicted to have emerged from the ribosome and be exposed to the cytosol. At increasing linker lengths, a progressive reduction in intensity in ^{1}H-^{15}N SOFAST HMQC spectra is observed, corresponding to the loss of disordered ddFLN5 structure. While folded resonances do not appear in the ^{1}H-^{15}N SOFAST spectra due to the reduced rotational correlation time of the folded domain (as discussed in Section 2.1.2), a selective labelling strategy based on perdeuteration and selective protonation of isoleucine side-chain methyl groups similar to that described in this thesis for TF^{ile} in Chapter 3 permitted the observation of the progressive population of the native ddFLN5 folded state on the ribosomes at linker length beyond 45 residues. The current understanding of the ddFLN5 RNCs thus provides an excellent framework for the study of the interaction of

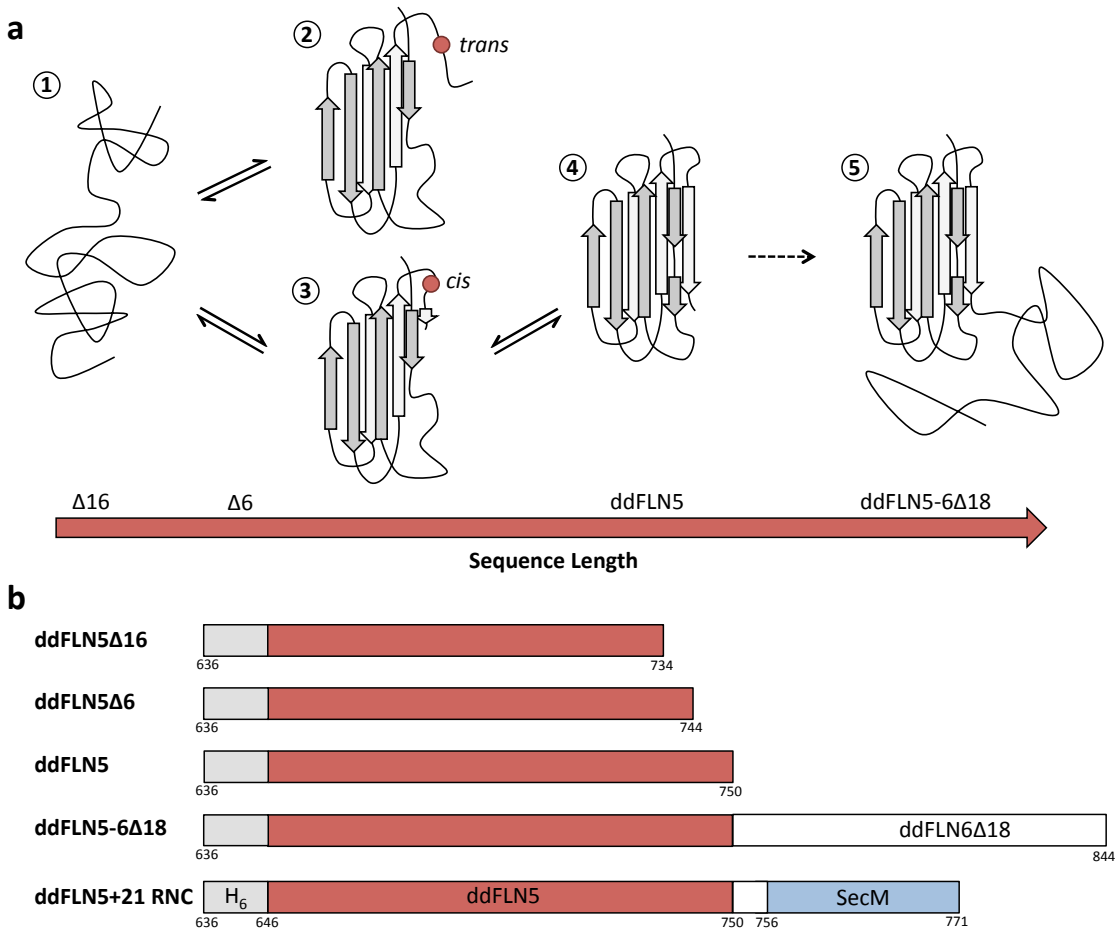

Figure 5.1: Overview of ddFLN5 truncation mutants and proposed co-translational folding pathway. (**a**) Schematic representation of the identified folding states populated by the various truncation mutants of ddFLN5. ddFLN5 constructs lacking 10 or more C-terminal residues (ddFLN5Δ10) do not adopt structure (1). The presence of residues pertaining to the last C-terminal β-strand in the native structure result in an equilibrium between the unfolded state and two partially folded states. The two folding intermediates (2/3) are related through proline isomerisation of P742. (**b**) ddFLN5 constructs used in the study detailed in this Chapter. All constructs contain a N-terminal hexa-histidine (H_6) tag for affinity purification (grey) and various lengths of the wild-type ddFLN5 sequence (red). ddFLN5-6 contains an additional 89 residues from the successive ddFLN6 domain (white) as example of sequential folding in multi-domain proteins. The ddFLN5+21RNC contains 6 linker residues and the SecM translational stalling sequence (blue) which site-specifically inhibits further translation and nascent chain release.

TF with nascent chains of various lengths.

In the study detailed in this Chapter, the focus lies on the shortest nascent chain length studied (ddFLN5+21 RNC) and the isolated reference unfolded state, ddFLN5Δ16 as reporters of the early interaction of TF with nascent chains.

5.2 Materials and Methods

Isolated ddFLN5, the various truncation mutants (ddFLN5Δ16, ddFLNΔ12, ddFLN5Δ6, ddFLN5-6) shown in Fig 5.1 and single tryptophan substitution mutants (Y655W, F665W, A668W, F672W and E724W) used throughout the study that follows (Section 5.5) and shown in Fig 5.8 have been prepared following the protocol outlined in Section 2.2.4.5. Purification of unlabelled 70S ribosomes proceeded as described in Section 2.2.5 with a notable difference being that final NMR buffers did not contain EDTA to prevent interference with the added Co^{2+} ions as described Section 5.6. The expression and purification of ^{15}N-labelled ddFLN5+21 RNCs is outlined in Section 5.4.1 and a detailed protocol can be found in Section 2.2.6.

5.3 The interaction of TF with isolated ddFLN5

In a parallel strategy to that described in Section 4.3 for the α-synuclein RNC, the interaction of TF with the ddFLN5 protein was first characterised in the absence of the ribosome. Various constructs sampling the folding intermediates of ddFLN5 described in Section 5.1 and Fig 5.1 were produced and probed for TF affinity by NMR as a basis for the observations made on the ribosome described in Section 5.4.

5.3.1 NMR characterisation of ddFLN5 isolated protein constructs

As was described in Section 5.1, the progressive truncation of C-terminal residues from ddFLN5 results in the gradual unfolding of the immunoglobulin fold transitioning through an intermediate species, distinguished by a non-native trans proline (P742). ddFLN5Δ16 is the shortest isolated protein construct used in this study, with 16 residues truncated from the full-length ddFLN5 sequence and representing a fully unfolded state across all studied temperatures. The ddFLN5Δ16 ^{1}H-^{15}N HSQC spectrum (100 μM protein at 10°C and pH 7.5) presents the narrow dispersion of resonances in the ^{1}H dimension (7.8-8.7 ppm) characteristic of an unfolded protein (Fig 5.2.a). This state resembles most closely the unfolded species detected for ddFLN5 RNC constructs used in this and previous studies (see Section 5.4).

The ddFLN5Δ6 truncation mutant (Fig 5.2.b) is observed to sample a mixture of states populated to similar extent. The resonances centred around 8.2 ppm overlay to a good

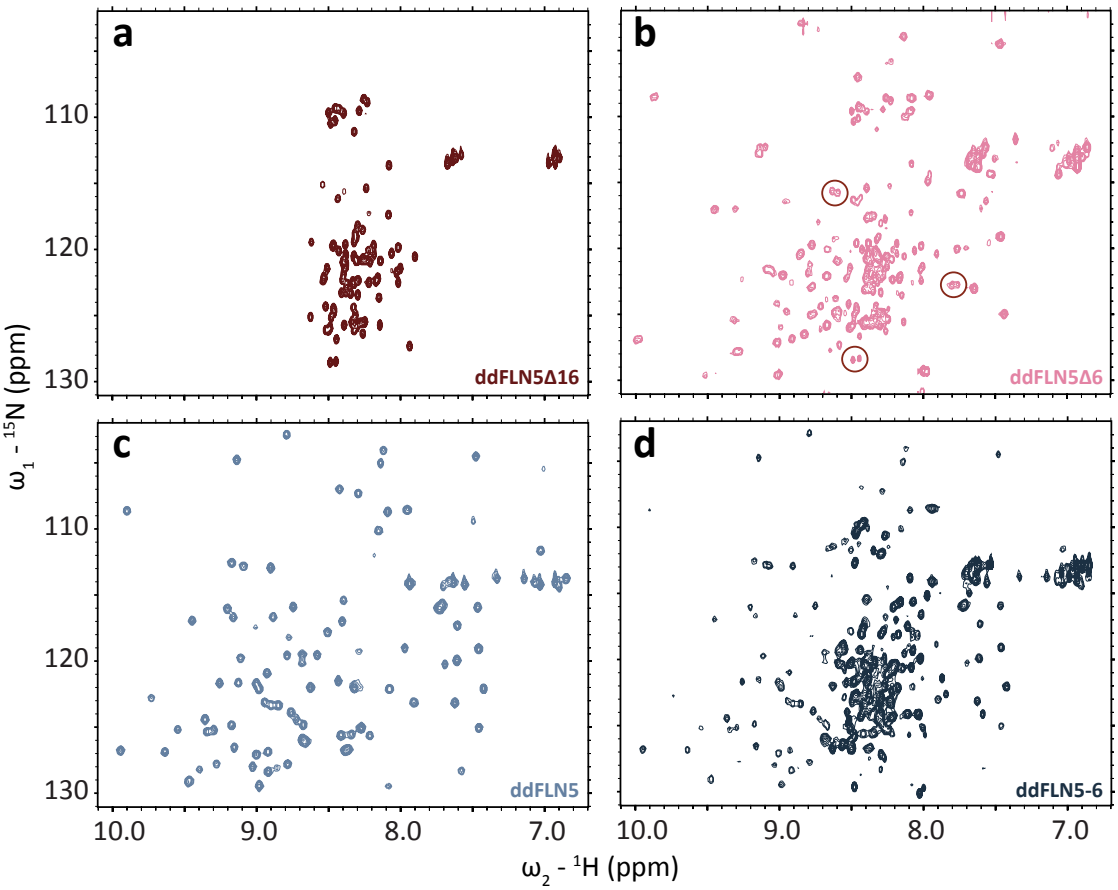

Figure 5.2: ^{1}H-^{15}N HSQC spectra of ddFLN5 constructs. ^{1}H-^{15}N HSQC spectra of the ^{15}N-labelled ddFLN5 constructs used in this study: the entirely unfolded ddFLN5Δ16 (100 μM) (**a**), ddFLN5Δ6 (50 μM) in equilibrium between the unfolded state and two distinct intermediate species (refer to Section 5.1) (**b**), natively folded ddFLN5 (100 μM). Pairs of split resonances are circled in red. (**c**) and ddFLN5-6Δ18 (100 μM) with a natively folded domain 5 and folding incompetent domain 6 (**d**) at 10°C and pH 7.5.

approximation with the ^{1}H-^{15}N HSQC resonances of ddFLN5Δ16, indicating that this species corresponds to the fully unfolded state. The well-dispersed resonances can readily be mapped to resonances in the native ddFLNL5 ^{1}H-^{15}N HSQC spectrum, indicating the presence of a native-like folding intermediate. Some but not all resonances are observed to split into two peaks (zoom in Fig 5.3.b). As discussed earlier, the detailed study by Waudby et al, (in preparation) revealed that the observed splitting of folded resonances corresponds to the population of two different folding intermediate species related through proline-isomerisation and associated structural rearrangements.

The full-length ddFLN5 (Fig 5.2.c) presents a well-dispersed ^{1}H-^{15}N HSQC spectrum characteristic of a folded protein. The ddFLN5-6Δ18 protein model contains 89 additional

residues from the sequential domain 6, which has been truncated to a folding incompetent state [258]. The isolated ddFLN5-6Δ18 approximates to the longest of ddFLN5 RNC constructs used, ddFLN5+110 RNC. ddFLN5-6Δ18 lacks only the SecM stalling sequence of this RNC construct. The ^{1}H-^{15}N HSQC spectrum of ddFLN5-6Δ18 (Fig 5.2.d) shows a fully folded domain 5 with unfolded domain 6 resonances centred around 8.2 ppm.

Moreover, all protein constructs in this Chapter contain a H_6-tag for affinity purification. The ddFLN5 constructs studied here are overall of a hydrophilic nature (Fig 5.8.a). No signature TF binding motifs, particularly enriched in aromatics and non-polar residues [159], as described in Section 1.2.2.3, are present in the sequence although two sites of above average hydrophobicity exist between residues E671-A676 and Y715-Y719 each containing two aromatic residues. Finally, since proline isomerisation is required for native folding of the ddFLN5 protein, P742 and the surrounding protein sequence might constitute a binding partner for the TF PPIase domain. The study of these protein constructs was therefore undertaken in order to establish a reference frame for the interpretation of the observations made in the case of the ddFLN5+21 RNC.

5.3.2 Characterisation of the interaction of TF with isolated ddFLN5

Equimolar amounts of TF were added to 100 μM ddFLN5Δ16, 50 μM ddFLN5Δ6, 100 μM ddFLN5 and 100 μM ddFLN5-6Δ18 in order to identify any pre-existing interaction between TF and the ddFLN5 substrates in the absence of additional factors such as the ribosome. No specific interaction was detected for any of the ddFLN5Δ16, ddFLN5Δ6 and ddFLN5 constructs upon equimolar addition of TF at both 10°C and 25°C (pH 7.5), as summarised in Fig 5.3. Addition of TF neither visually altered the appearance of the ddFLN5Δ16, ddFLNΔ6 and ddFLN5 ^{1}H-^{15}N HSQC spectra, nor did resonance intensity analysis display specific sites of enhanced broadening that would be indicative of interaction. Moreover, TF did not specifically interact with the ddFLN5Δ6 sub-population that presents a non-native *trans* proline, nor was it found to alter the equilibrium between *cis* and *trans* states.

5.3.2.1 Characterisation of the interaction of TF with ddFLN5-6Δ18

Addition of equimolar TF to 100 μM ^{15}N-labelled ddFLN5-6Δ18 at 25°C did not result in notable changes in the ^{1}H-^{15}N HSQC spectrum and no effect on resonance intensities was detected (Fig 5.4.b). At 10°C however, a significant increase in resonance intensities of domain 5 (residues 646-750) is detected to 1.3 fold the original intensity (Fig 5.4.c). This is at odds with the expected selective broadening of resonances upon interaction with TF.

Resonances originating from the unfolded domain 6 display more variable behaviour. A detailed description of the resonance intensities is hindered by the incomplete assignment

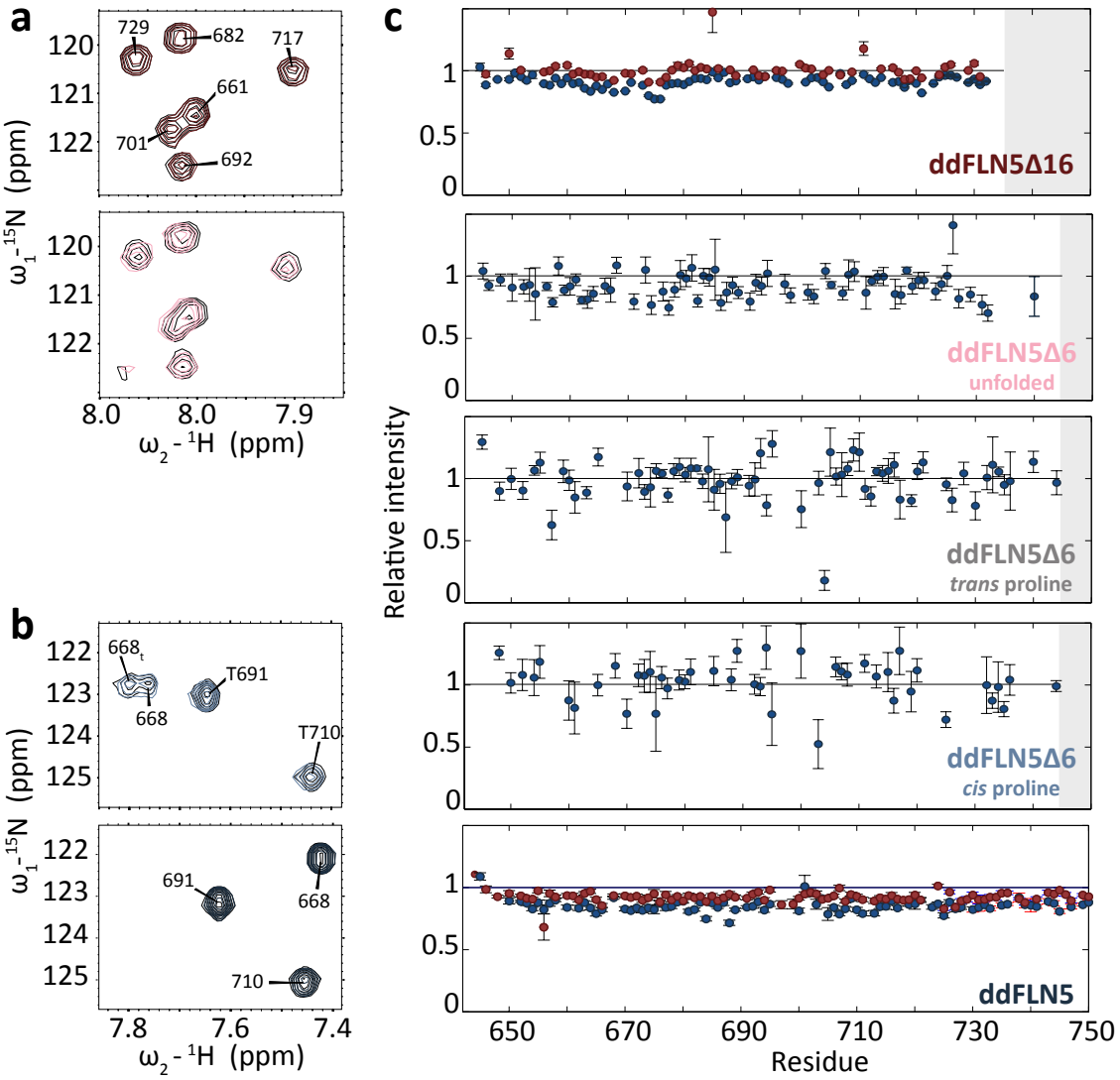

Figure 5.3: Interaction of TF with ddFLN5 truncation mutants. (**a**) Select ^{1}H-^{15}N HSQC unfolded resonances of ddFLN5 truncation mutants ddFLN5Δ16 (red) and ddFLN5Δ6 (pink) in the presence and absence (colour/black) of equimolar TF at 10°C. (**b**) Select ^{1}H-^{15}N HSQC folded resonances of ddFLN5 truncation mutants ddFLN5Δ6 (light blue) and ddFLN5 (dark blue) in the presence and absence (colour/black) of equimolar TF at 10°C. Resonances assigned to the trans proline state of ddFLN5Δ6 are marked 'T'. (**c**) Relative ^{1}H-^{15}N HSQC resonance intensities of ddFLN5Δ16, the three states of ddFLN5Δ6 and ddFLN5 in the presence of equimolar TF at 10°C (blue) and 25°C (red). Grey shaded regions correspond to residues absent in the particular construct.

of resonances in this region of the ddFLN5-6Δ18 sequence due to spectral overlap and limited sample stability (Anaïs Cassaignau, UCL). From the available data it appears that some resonances are reduced to 60% of their initial intensity in the presence of TF. The affected residues are 758 to 767 and a second, stronger site centred at N799. The first site is enriched in phenylalanine (2) and serine (3) residues, while the second site displays a number of negatively charged residues (4).

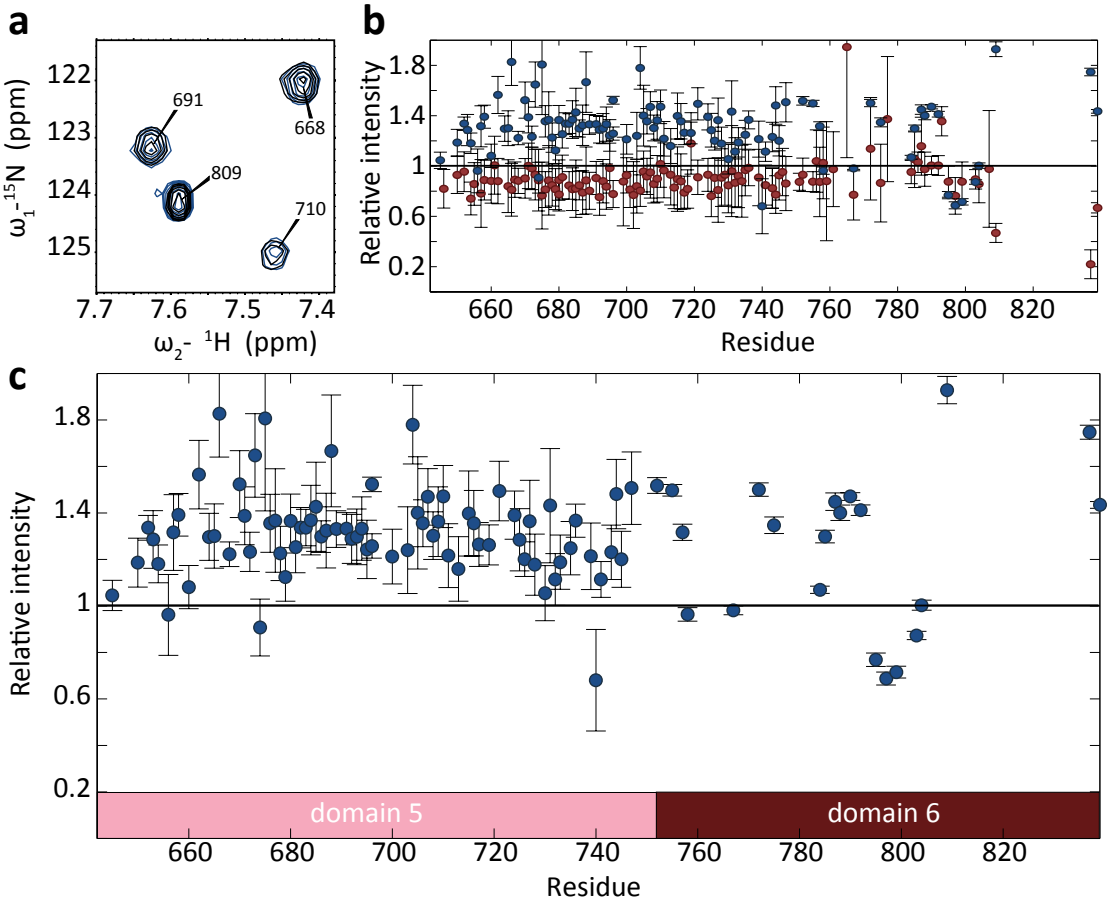

Figure 5.4: Interaction of TF with ddFLN5-6Δ18. (**a**) Region of the ^{1}H-^{15}N HSQC spectra of 100μM ^{15}N-labelled ddFLN5-6Δ18 in the absence (black) and presence of equimolar TF (blue) at 10°C and pH 7.5. (**b**) Relative ^{1}H-^{15}N HSQC resonance intensities upon addition of equimolar TF at 10°C (blue) and 25°C (red). (**c**) Relative ^{1}H-^{15}N HSQC ddFLN5-6Δ18 resonance intensities upon addition of equimolar TF at 10°C as in **b** reveal overall increase in intensity of the detected NMR signals in the presence of TF, with selective broadening of residues 758-767 and 790-810 in the ddFLN6 linker sequence.

5.3.3 Concluding remarks

The study of the interaction of TF with the isolated ddFLN5 truncation mutants detailed above demonstrates that TF shows no inherent affinity for the folding intermediates populated by ddFLN5 *in vitro* in the absence of ribosomes. The agreement between the absence of extended stretches of above average hydrophobicity and the absence of detected interaction with TF reaffirms the predictive power of amino acid sequence analysis to identify TF substrates as described in Section 4.1.1.

TF however does appear to display affinity for the ddFLN6 linker that is exposed in longer ddFLN5 RNC constructs at 10°C. Under our established 'typical' experimental conditions for the study of nascent chains by NMR (10°C for unfolded, and 25°C for folded nascent chains), TF does however not interact selectively with ddFLN5 in the absence of the ribosome. The ddFLN5 nascent chain thus provides a complementary null system for the study of the TF interaction with emerging nascent chains on the ribosome to the previously described α-synuclein RNC.

5.4 The interaction of TF with ddFLN5+21 RNC

The ddFLN5+21RNC sequence is composed of the full-length ddFLN5, followed by a stalling sequence from the regulatory protein SecM which arrests nascent chain production on the ribosome (see Fig 5.1.b). Four residual cloning residues serve as a spacer between the folding competent ddFLN5 domain and the SecM translational arrest sequence, distancing the ddFLN5 domain at 21 residues from the PTC. The full sequence is included in Appendix A.1.3. With approximately 30 residues required to bridge the ribosomal exit tunnel, at least some of the C-terminal strand G of ddFLN5 is predicted to be (partially) buried inside the ribosomal exit tunnel, preventing native folding of the domain. In accord with this, the ddFLN5+21 RNC was found to display predominantly non-dispersed cross-peaks in SOFAST HMQC spectra at 25°C, with chemical shifts closely similar to those of the isolated ddFLN5Δ12 [65, 258].

In this Section, the ddFLN5+21 RNC was used as a model system to characterise the interaction of TF with an unfolded nascent chain. The preparation of the ddFLN5+21 RNC is briefly outlined followed by a detailed characterisation of the ddFLN5+21 RNC at 10°C. Then the effect of TF presence on the behaviour of the nascent chain is discussed.

5.4.1 Preparation of the ddFLN5+21 RNC

The detailed protocol for the expression and purification of uniformly ^{15}N-labelled ddFLN5+21 RNC can be found Section 2.2.7. In short, transformed *E. coli* BL21*(DE3) Gold cells are grown in unlabelled growth medium to high optical densities ($OD_{600} \sim 4$). Cells are

then transferred to isotope enriched medium for a short induction period (30 min). This ensures that the majority of ribosomes are NMR silent while the ^{15}N isotope is selectively incorporated into the nascent chain during expression, for minimal background signal from the flexible L7/L12 stalk region of the ribosome in the ^{1}H-^{15}N SOFAST HMQC spectrum.

Ribosomes are isolated from the lysate by sucrose cushion centrifugation. The ribosomal fraction (20 nmol per litre culture) is enriched in nascent chain occupancy by affinity chromatography for the nascent chain H_6-tag. Finally a 10-35% sucrose gradient centrifugation step allows the separation of 70S RNC from contaminant 30S and 50S ribosomal subunits. A typical preparation yields 3-6 nmol of RNC material per litre of bacterial culture, yielding 1-2 NMR samples at maximal ribosome concentrations (10 μM).

5.4.2 Characterisation of the ddFLN5+21 RNC by NMR

While previous characterisation of the ddFLN5+21 RNC by NMR [65] has been undertaken at 25°C, here, the choice was made to investigate the behaviour of the ddFLN5+21 RNC at 10°C for the decreased contribution of the background L7/L12 signal (relative to 25°C, Fig 5.5.c) and reduced resonance linewidths for the nascent chain due to decreased solvent exchange.

The ^{1}H-^{15}N SOFAST HMQC spectrum of 10 μM ddFLN5+21 RNC at 10°C, pH 7.5, displays the expected narrow resonance dispersion with all backbone amide signals between 8-8.6 ppm. 46 distinct resonances for a total of 132 residues are detected with a signal-to-noise of ∼20 achieved within 1 hour of acquisition.

The resolution of the NMR spectrum was increased relative to the cited study at 25°C by doubling the number of increments in the indirect dimension at the expense of the number of scans acquired, reducing the S/N per unit time by $\sqrt{2}$. The relationship between the various NMR parameters is discussed in Section 2.1.1. The resulting increased resolution reduced the spectral overlap of the broad RNC signals and allowed the more accurate determination of the chemical shifts of HMQC resonances. As shown in Fig 5.5.b, the ddFLN5+21 RNC SOFAST HMQC spectrum overlays well with the unfolded ddFLN5Δ16 reference spectrum allowing for ready transfer of resonance assignment.

ddFLN5+21 RNC resonance assignments were ranked into 5 categories: 1) resonances in good agreement with ddFLN5Δ16 chemical shifts (differences smaller than 0.005 ppm and 0.05 ppm in the ^{1}H and ^{15}N dimensions respectively), 2) resonances that could be readily identified but that overlapped with 70S background signals (Fig 5.5.c) potentially leading to distortions of the baseline and thus errors in the resonance intensities measured, 3) resonances undergoing chemical shift perturbations relative to the ddFLN5Δ16 and thus could not be assigned with great certainty, 4) overlapped resonances and finally 5) ddFLN5Δ16 resonances visibly absent from the ddFLN5+21 RNC spectrum were

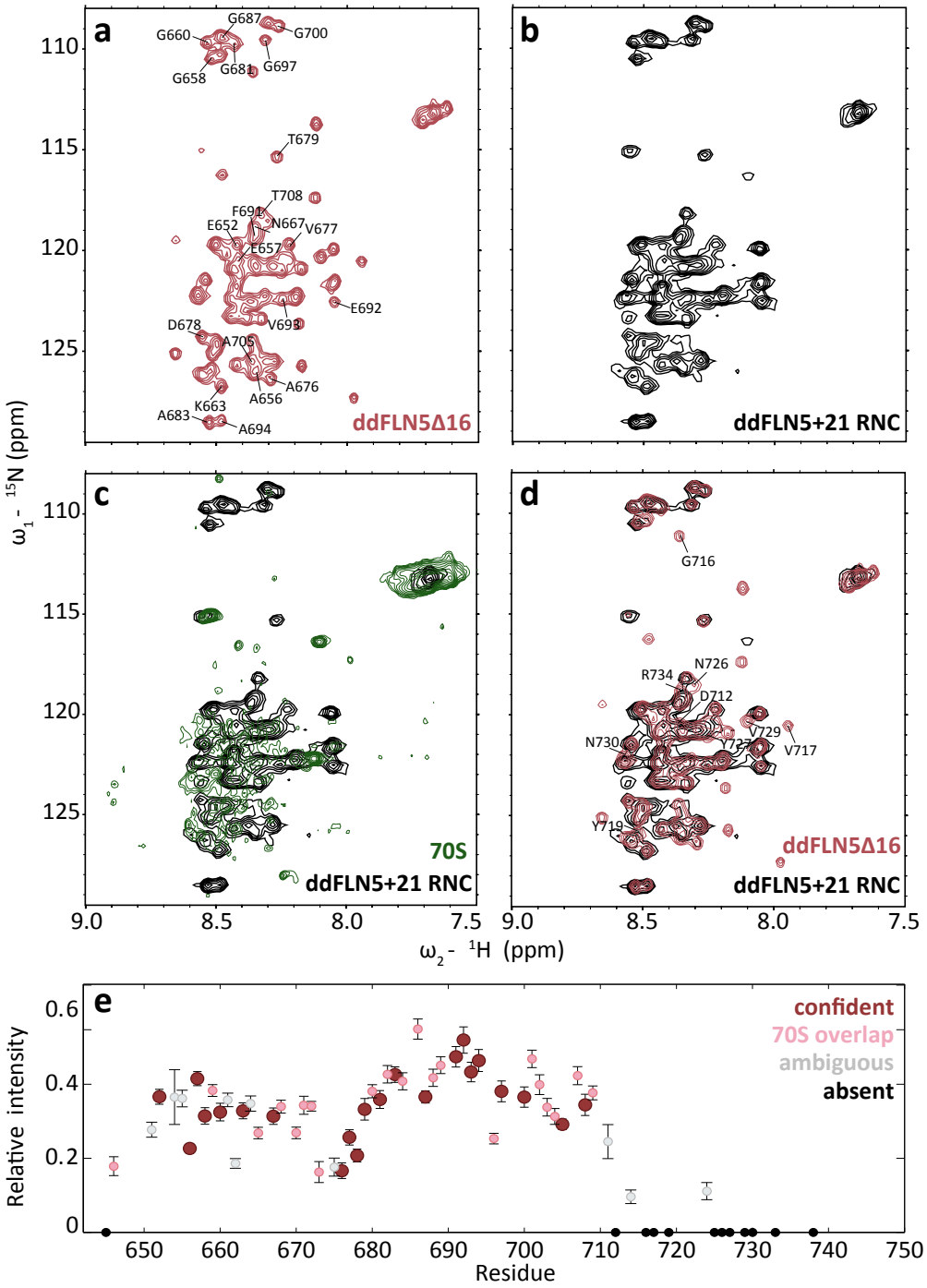

Figure 5.5: Characterisation of the ddFLN5+21 RNC at 10°C by NMR (**a**) ^{1}H-^{15}N SOFAST-HMQC of 10μM ddFLN5Δ16 (pink). Resonances confidently assigned in the ddFLN5+21 RNC are annotated. (**b**) ^{1}H-^{15}N SOFAST-HMQC of 10μM ddFLN5+21 RNC (black). (**c**) ^{1}H-^{15}N SOFAST-HMQC of ddFLN5+21 RNC (black, as in b) overlaid with ^{1}H-^{15}N SOFAST-HMQC of the 70S (green). (**d**) ^{1}H-^{15}N SOFAST-HMQC of ddFLN5+21 RNC (black, as in b) overlaid with ^{1}H-^{15}N SOFAST-HMQC of ddFLN5Δ16 (pink, as in a). (**d**) ddFLN5+21 RNC resonance intensities relative to ddFLN5Δ16. Individual resonances are classified according to confidence level in assignment and peak height : highest confidence in red, resonances with minor overlap with 70S resonances in pink, ambiguous assignments in grey and absent resonances in black.

identified and classed separately. Overlapped resonances (category 4) were excluded from further analysis since signal overlap results in baseline distortions hindering an accurate intensity analysis and prevents the certain transfer of resonance assignments.

Observable, and assigned, resonances are on average broadened to 35% of the signal intensity of 10 μM ddFLN5Δ16 in isolation. A sequence dependent broadening pattern emerges as shown in Fig 5.5.e. Resonances originating from N-terminal residues 650 to 670 are uniformly broadened to 35%. Subsequent residues 672 to 680 are reduced to 20 %. From residue 680 to 695, resonances show 40% intensity slowly decreasing to 35% at 710. From residue 710 onwards, no further resonances are detected. Fig 5.5.d highlights the absence of these resonances in the ddFLN5+21 RNC spectrum. Extreme N-terminal residues 646-650 further appear reduced in intensity relative to the subsequent regions.

The ddFLN5+21 RNC presents a ^{1}H-^{15}N SOFAST HMQC of remarkable quality as association with the ribosome particle could potentially lead to extremely rapid relaxation (R_2 >1000 s^{-1}) and thus an unobservable state. The behaviour of ddFLN5+21 RNC thus appears to mirror the conformational freedom of the dynamic L7/L12 regions, rotating independently of the ribosome.

The specific broadening at the N-terminus and residues 672-680 may correspond to sites of transient interaction with the ribosome surface or local conformational exchange specific to the nascent chain. The effect of chemical exchange on resonance intensities is described in Section 2.1.6.1. C-terminal resonances from residues 712 onwards (residues 771 is the last residue in the ddFLN5+21 RNC sequence) are not present in the ddFLN5+21 RNC spectrum. The limited motional freedom of $\sim$30 residues inside the ribosome tunnel readily explains their broadening beyond detection. A gradual recovery of signal intensities would be expected as flexibility is gained progressively as residues are further removed from the ribosome surface. Instead, a rapid transition in signal from 35% to 0% is observed between residue 709 and 712. It is therefore likely that the origin of the broadening observed in the last 60 residues can be found in an exchange event on the timescale of the NMR experiment in the form of interaction with the ribosome surface or conformational exchange.

The experimental set-up employed here for the NMR study of the ddFLN5+21 RNC at 10°C with increased spectral resolution, has permitted a residue specific analysis of the dynamic behaviour of the ddFLN5+21 nascent chain on the ribosome. The high quality of the NMR spectrum provides ideal conditions for the characterisation of the TF with the ddFLN5+21 RNC.

5.4.3 Determining the attachment of the ddFLN5+21 nascent chain to the ribosome

A stringent protocol based on both NMR and biochemical experiments for the verification of attachment of the nascent chain to the ribosome has been established in our laboratory (Helene Launay, doctoral thesis, 2011, and Cassaignau *et al.*, in preparation) in order to avoid the interpretation of the NMR data in the presence of released protein. The reduced rotational correlation time of the released protein typically results in undesirable significant contribution to the NMR spectrum. The effect of the rotational correlation time (τ_c) on NMR signal intensity is described Section 2.1.2.

^{15}N diffusion : ^{15}N SORDID diffusion experiments of the ddFLN5+21 RNC were recorded at two gradient strengths: 10.4% and 69.5% G_{max} as described in Section 2.3.2.5. The one-dimensional ^{15}N-edited spectra at the start of acquisition are shown in Fig 5.6.c. The diffusion coefficients of the nascent chain over time can be calculated from the ratio of the signal integral from equation 2.1.16. At early time-points, the diffusion coefficient was found to be $1.2 \pm 0.3 \times 10^{-11}$ m^2.s^{-1} in good agreement with experimental diffusion coefficients determined for the 70S particle (1×10^{-11} m^2.s^{-1} and $1.85 \pm 0.15 \times 10^{-11}$ m^2.s^{-1} at 4°C and 25°C respectively (Helene Launay, doctoral thesis, 2011)). Release of the nascent chain is reflected by an increase in the measured diffusion coefficient towards the isolated protein value of $1.3 \pm 0.05 \times 10^{-10}$ m^2.s^{-1} (Marilia Karyadi, doctoral thesis, 2012). Fig 5.6.d shows the progressive increase of the experimental diffusion coefficient for the ddFLN5+21 RNC over the course of multiple days. At t=18 hr the detected change in the diffusion coefficient exceeds the experimental error of the value at t=0 hr which is typically used as a cut-off for data accumulation on the RNC. The change in the dissociation constant over the course of data acquisition corresponds to only a fraction of released nascent chains relative to the attached population, as the shorter rotational correlation time of the released nascent chain results in a large fractional contribution to the NMR signal.

Ribosomal background labelling : ^{15}N background labelling of the ribosome was determined from the intensity ratio of a ^{15}N-filtered and ^{15}N-edited set of one-dimensional NMR experiments described in Section 2.3.2.1. While the ^{15}N-filtered experiment (Fig 5.6.e) reflects the fraction of NMR silent (^{14}N) ribosomes, the ^{15}N edited spectrum reports on a combination of the nascent chain and ^{15}N labelled ribosome. The intensity ratio of the two spectra between 8.5 and 9 ppm, where only the folded ribosomal protein signals are expected to be observed, was found to be 0.2, which translates to 15% background labelling of the ribosomes present in the NMR sample. The major contribution to the ^{15}N signal detected in the SORDID diffusion experiments described above is therefore arising from

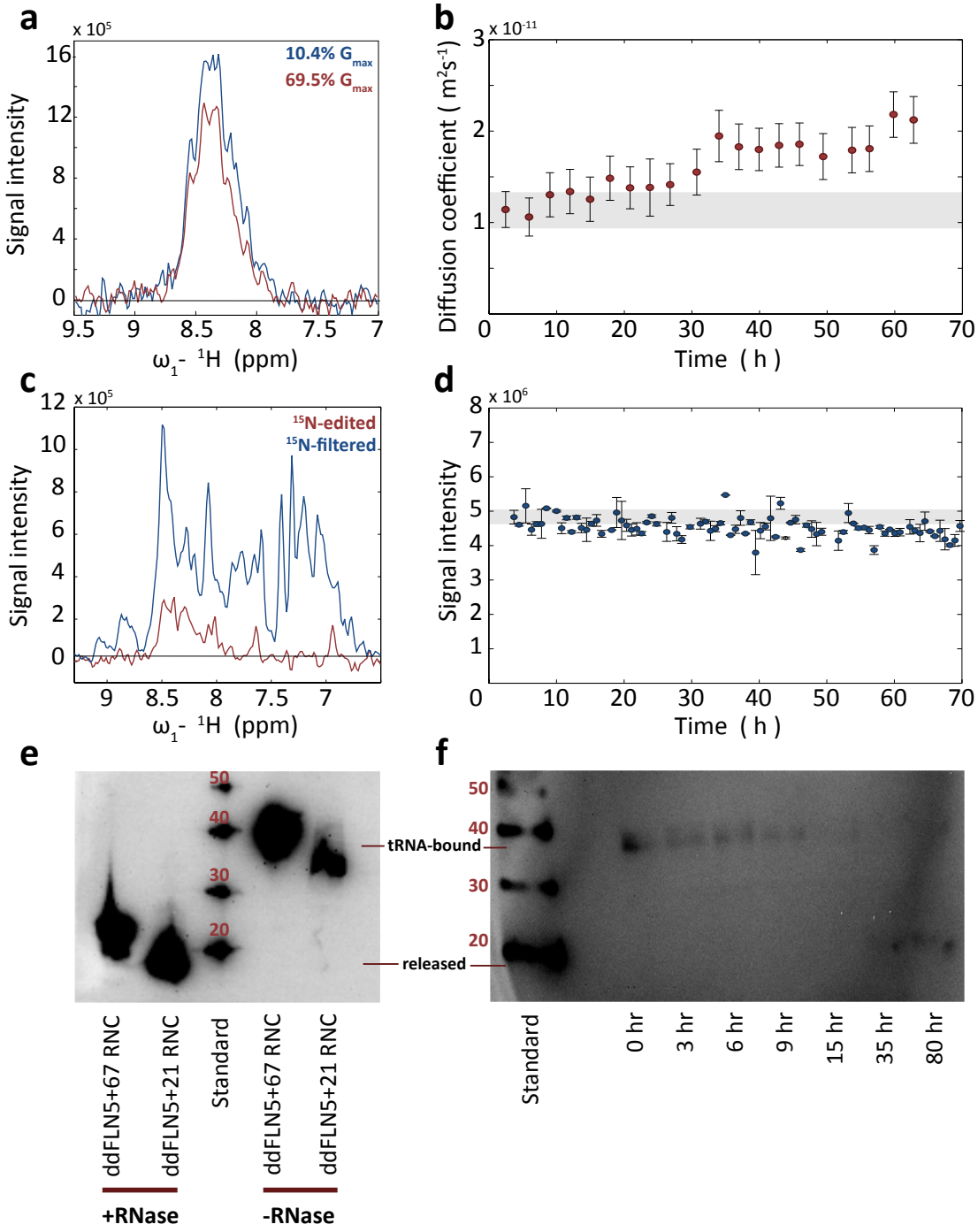

Figure 5.6: Stability of the ddFLN5+21 RNC at 10°C. (**a**) ^{15}N SORDID diffusion 1D spectra of the ddFLN5+21 RNC at 10.4% and 69.5% maximum gradient strength. (**b**) Calculated diffusion coefficients for the ddFLN5+21 nascent chain over time. (**c**) ^{15}N-edited and ^{15}N-filtered 1D spectra of ddFLN5+21 RNC. (**d**) ^{1}H-^{15}N SOFAST HMQC first increment signal integral of the ddFLN5+21 nascent chain over time. All errors were determined by integration of an identical sized region of the noise in the NMR spectra. (**e**) Anti-his detected western blot of 4 pmol prepared ddFLN5+21 RNC and a ddFLN5+67 RNC control in the absence and presence of RNase. (**f**) Anti-his detected western blot of ddFLN5+21 RNC stability over time.

the nascent chain, indicating that the time-dependent increase of the diffusion coefficient indeed reports specifically on nascent chain attachment to the ribosome.

^{1}H-^{15}N SOFAST HMQC signal integrals : Finally, changes over time in the signal intensities in the first increment of 2D experiments such as the ^{1}H-^{15}N SOFAST HMQC report on the stability of the nascent chain and flexible ribosome regions (L7/L12 and S1) in terms of degradation and release. The variation in the signal integrals between 7.8 and 8.7 ppm over time are shown in Fig 5.6.f. Significant reduction of the SOFAST HMQC signal coincided with the observed changes in the diffusion coefficient at 18 hours.

Biochemistry : Nascent chain attachment can be readily determined biochemically as the covalent bond of the SecM stalled nascent chain to the peptidyl tRNA results in an upshift of 16 kDa of the nascent chain band relative to the released protein on an SDS-PAGE gel. The covalent bond, prone to hydrolysis, can be maintained under low pH conditions (pH 5.8) as shown in Fig 5.6.e. The fractional changes in the tRNA-bound nascent chain were monitored over the course of the NMR experiment by Anti-His detected western blot immuno-assay (Fig 5.6.f). The time-course of the ddFLN5+21 RNC demonstrates that the ribosome-nascent chain complex is stable for a minimum of 9 hours, as indicated by the appearance of the released species at t = 35hr.

The combination of the three distinct experimental reporters of nascent chain attachment (SDS-PAGE nascent chain band up-shift, ^{15}N diffusion and SOFAST HMQC signal integrals) indicate the ddFLN5+21 RNC is stable over a period of 9-18 hours at 10°C. The findings described in Section 5.4.2 and the study that follows report on NMR data acquired in the first 8 hours.

5.4.4 NMR investigation of the interaction of ddFLN5+21 RNC with TF

The addition of TF to 10 μM ^{15}N-labeled ddFLN5+21 RNC at 10°C and pH 7.5 results in broadening of ^{1}H-^{15}N SOFAST HMQC resonances. No notable changes in resonance chemical shift positions are detected (Fig 5.7.a and b). Resonances were classified into the 5 categories described above for the analysis of their relative intensities in the presence of TF.

At one molar equivalence TF (Fig 5.7.c), ddFLN5+21 RNC resonances appear uniformly broadened to 75% of their intensity in the absence of TF. No selective broadening pattern beyond the error on the resonance intensity measurements were observed. Furthermore, no new ^{1}H-^{15}N SOFAST HMQC cross-peaks are detected in the comparison with the NMR spectrum before addition of TF, indicating the C-terminal residues, from 712 onwards, remain broadened beyond detection.

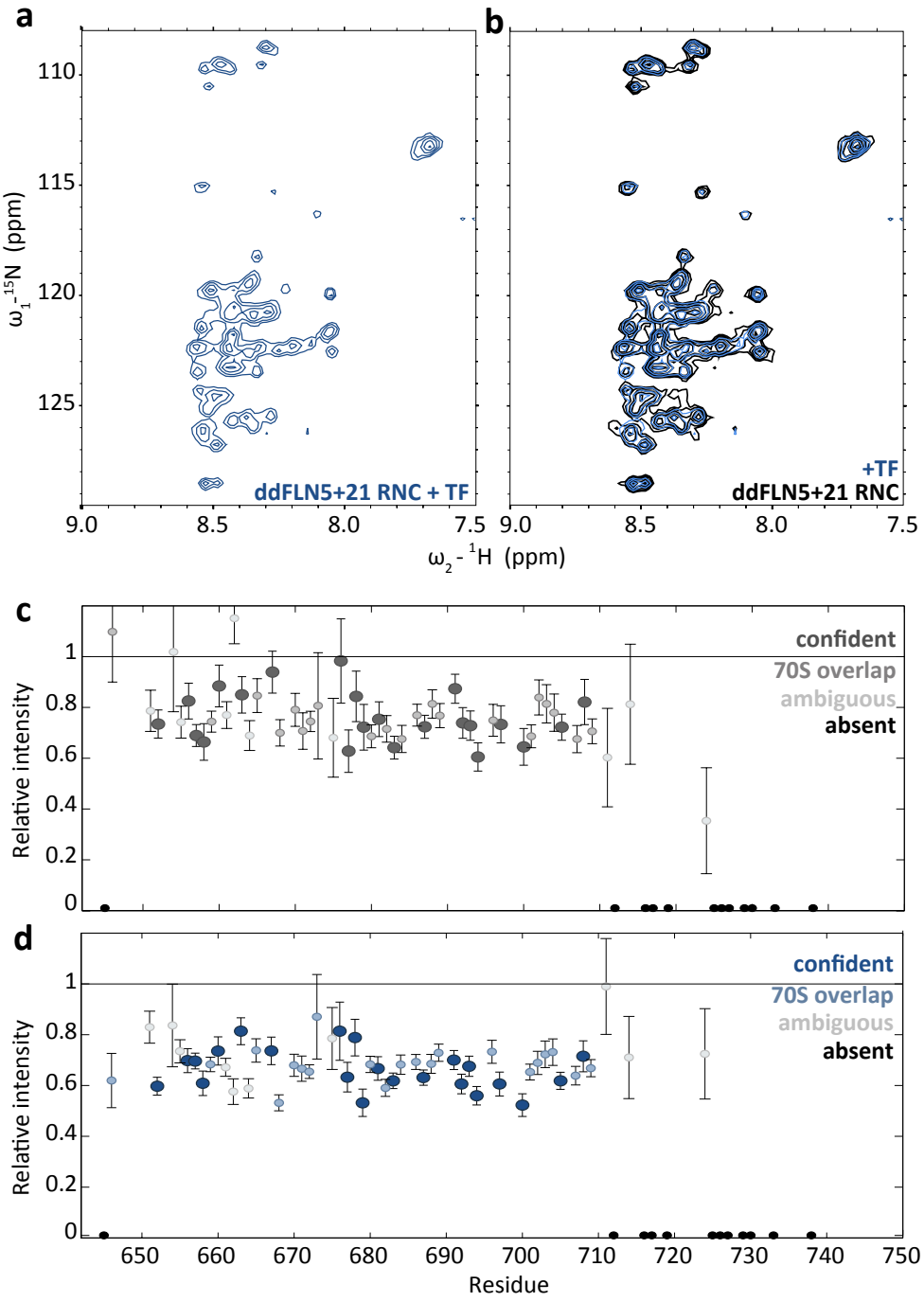

Figure 5.7: TF interaction with the ddFLN5+21 RNC. (**a**) ^{1}H-^{15}N SOFAST HMQC spectrum of ddFLN5+21 RNC in the presence of 3 molar equivalence TF. (**b**) Overlay of the ^{1}H-^{15}N SOFAST HMQC spectra of ddFLN5+21 RNC in the absence (black) and presence of 3 molar equivalence TF (blue). (**c**) Relative ddFLN5+21 RNC resonance intensities in the presence of 1 equivalence TF. (**d**) Relative ddFLN5+21 RNC resonance intensities in the presence of 3 equivalence TF. Individual resonances are classified according to confidence level in assignment and peak height : highest confidence in dark blue, resonances with minor overlap with 70S resonances in blue, ambiguous assignments in grey and absent resonances in black.

At three molar equivalence of added TF (Fig 5.7.d), close to saturation of the TF interaction as predicted by the mathematical model of the complex TF equilibrium developed in Section 3.4, the relative intensity drop further down to 65% of the initial intensities in the absence of TF.

TF does not appear to interact specifically with regions of the ddFLN5+21 RNC. The presence of TF at the exit tunnel does however affect the conformational freedom of the nascent chain resulting in uniform broadening. This reduced conformational space may be conducive to protein folding or prevents non-desired interactions with cytosolic factors.

5.4.5 Concluding remarks

The investigation of the ddFLN5+21 RNC at 10°C has permitted the acquisition of high resolution NMR spectra The resonance intensities across the RNC could be compared at near residue specific level and variation in the extent of broadening across the sequence where observed. This indicates that the presence of the ribosome through tethering of the nascent chain and through potential interactions with the ribosome surface affects the behaviour of the ddFLN5+21 RNC relative to the unfolded ddFLN5Δ16. Although the specific differences could be localised to exact regions of the protein, the precise nature of the effect remains unknown.

The presence of TF at the ribosomal exit tunnel impacts on the conformational space of the ddFLN5+21 RNC in a concentration dependent manner. At saturation point, the presence of TF causes broadening of the ddFLN5+21 RNC resonances to 65%. The presence of TF appears to perturb the dynamics of the ddFLN5 nascent chain uniformly across the sequence. The results presented here suggest TF does not specifically interact with a particular region of the ddFLN5+21 nascent chain but has highlighted the capacity of NMR to investigate the TF interaction with the nascent chain at a residue specific level.

5.5 The interaction of TF with single tryptophan substitution ddFLN5Δ16 mutants

The study described in the previous Sections 5.3 and 5.4 focussed on the TF interaction with unfolded ddFLN5 in isolation and on the ribosome. The multi-domain ABP120 protein from which the ddFLN5 is a sub-domain, is not a native substrate of TF and does not appear to recruit TF *in vitro*. In Chapter 4, a highly hydrophobic sequence was substituted into the native α-synuclein sequence to enhance TF affinity. In a similar vein, we here develop a strategy to recruit TF to the ddFLN5 constructs by single amino acid substitution in order to complement our current picture of the TF interaction with protein substrates on the ribosome by including a scenario where a specific interaction site is present.

5.5.1 Strategy

The strategy pursued here is based on single non-destabilising amino acid substitutions. It was found previously that TF displayed a particular preference for aromatic residues, notably interacting with each of three tryptophan residues present in the PhoA sequence [130, 159]. The native ddFLN5Δ16 sequence presents two sites of above average hydrophobicity: E671-A676 and Y715 -Y719, both containing two aromatic residues each (Fig 5.8.a). Single amino acid substitutions to Trp residues were made around these sites in order to enhance the local hydrophobicity and to create a more favourable site for TF interaction. Residues were selected based on conservation and ensured to be non-destabilising to the native ddFLN5 fold as predicted by the PoPmusic server [259] in the attempt to produce folding competent full-length ddFLN5 mutant products.

Five sites were selected for mutagenesis: Y655, F665, A668, F672 and E724 (Fig 5.8.b). Y655 is in close structural proximity of the E671-A676 site of existing hydrophobicity permitting the formation of a three dimensional hydrophobic patch in the ddFLN5 folded state (Fig 5.8.e). The tryptophan substitution here is predicted to be marginally destabilising ($\Delta\Delta$G of 0.31) while it does not alter the mean hydrophobicity of the protein sequence (Fig 5.8.a). F665 and A668 ($\Delta\Delta$G of 0.24 and -0.12 respectively) extend the existing hydrophobic site between E671 and A676 with the A668W mutants resulting in a small increase in local hydrophobicity. Mutation of F672 does not alter the hydrophobicity or length of this ddFLN5 region but alters its sequence composition and is predicted to destabilise the native ddFLN5 structure ($\Delta\Delta$G of 0.99). The final site of mutation, E724 ($\Delta\Delta$G of -0.13) increases the mean hydrophobicity of the sequence (Fig 5.8.a) extending the second site of pre-existing hydrophobicity in the ddFLN5 sequence at residues Y715-Y719.

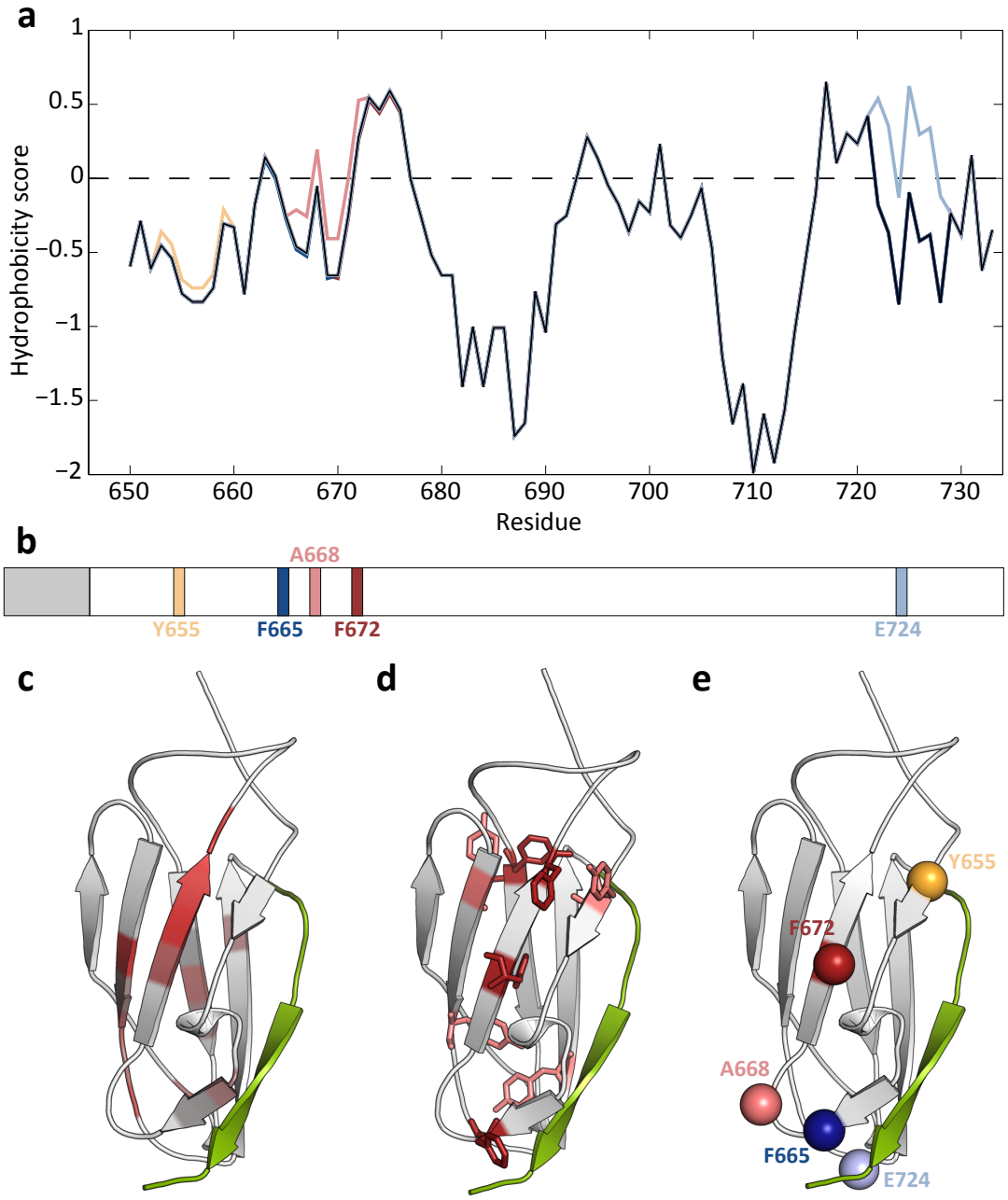

Figure 5.8: ddFLN5Δ16 tryptophan mutants. (**a**) Roseman hydrophobicity plots of ddFLN5Δ16 and five single tryptophan substitutions. ddFLN5Δ16 presents two sites of relative high hydrophobicity around E671-A676 and Y715 -Y719. The tryptophan substitutions do not significantly alter the hydrophilic nature of ddFLN5Δ16. (**b**) Localisation of the engineered tryptophan substitutions on the ddFLN5Δ16 sequence. (**c**) Mapping of sites of hydrophobicity on the folded ddFLN5 structure. Truncated C-terminal residues in ddFLN5Δ16 in green. (**d**) Native aromatic residues on the ddFLN5 structure: tyrosine in pink, phenylalanine in red. (**e**) Sites of single amino acid substitutions mapped on the ddFLN5 structure. Colours as in a and b.

5.5.2 NMR characterisation of ddFLNΔ16 tryptophan mutants

The five tryptophan substitution mutant proteins were successfully sub-cloned, expressed and purified as described in Sections 2.2.1, 2.2.3 and 2.2.4.5. The ^{1}H-^{15}N HSQC spectra of 80 μM F665W and E724W at 500 MHz and 10°C display the characteristic poor resonance dispersion of a disordered protein. Comparison to a ^{1}H-^{15}N HSQC spectrum of 100 μM ddFLN5Δ16 acquired under identical conditions shows that these spectra largely overlay with small chemical shift disruptions in resonances corresponding to residues in close sequence proximity to the site of mutation (Fig 5.9.a). The resonance intensities fluctuate randomly around their WT ddFLN5Δ16 values with no clear pattern discernible based on sequence proximity or composition. The detected intensity fluctuations do also not appear to correlate with any structural elements of the native ddFLN5 (Fig 5.10.a and b).

The ^{1}H-^{15}N SOFAST HMQC spectra of 100 μM of tryptophan mutants Y655W, A668W and F672W show similarly non-dispersed cross-peaks at 700 MHz and 10°C (Fig 5.9). Chemical shift changes from the WT ddFLN5Δ16 are more pronounced than in the previous two mutants and transfer of the existing ddFLN5Δ16 assignment was non trivial, significantly reducing the number of probes available in these substrates. In this instance, a clear trend can be observed from the resonance intensity profiles relative to the WT ddFLN5Δ16 construct. While the N-terminus presents similar random intensity variation as detected for the F665W and E724W ddFLN5Δ16 mutants, intensities around residue K680 appear conserved while further C-terminal residues are broadened to ∼40% in all three tryptophan mutants. The detected broadening mainly maps to what corresponds to β-strands CDEF in the native ddFLN5 structure (Fig 5.10.b).

Moreover a limited number (20) of well-dispersed cross-peaks were detected in the ^{1}H-^{15}N SOFAST HMQC of Y655W, A668W and F672W proteins, spread over 8.7-10 ppm in the ^{1}H dimension suggesting the presence of a folded species. Similar non native-like disperse resonances have been previously observed in ddFLN5Δ12 ^{1}H-^{15}N SOFAST HMQC spectra, a longer ddFLN5 truncation product which does not yet present the folding intermediate states associated with proline isomerisation as described Section 5.1 (Fig 5.1.a). The observed disperse peaks in the tryptophan mutants overlay reasonably well with the disperse peaks in the ddFLN5Δ12 SOFAST spectrum (Fig 5.10.e) but ddFLN5Δ12 present more disperse resonances ($\sim$ 60) with significantly lower intensities: 7-8 fold weaker than in the tryptophan mutants. The similarities suggest a common origin to the dispersed peaks in the tryptophan mutants and ddFLN5Δ12. While the correlation of C-terminal broadening and the appearance of a novel set of resonances in the Y665W, A668W and F672W ^{1}H-^{15}N SOFAST HMQC suggests the tryptophan mutant proteins might be undergoing conformational exchange, an equally likely explanation is the co-purification of a small folded protein contaminant with higher affinity for the inserted tryptophan residues.

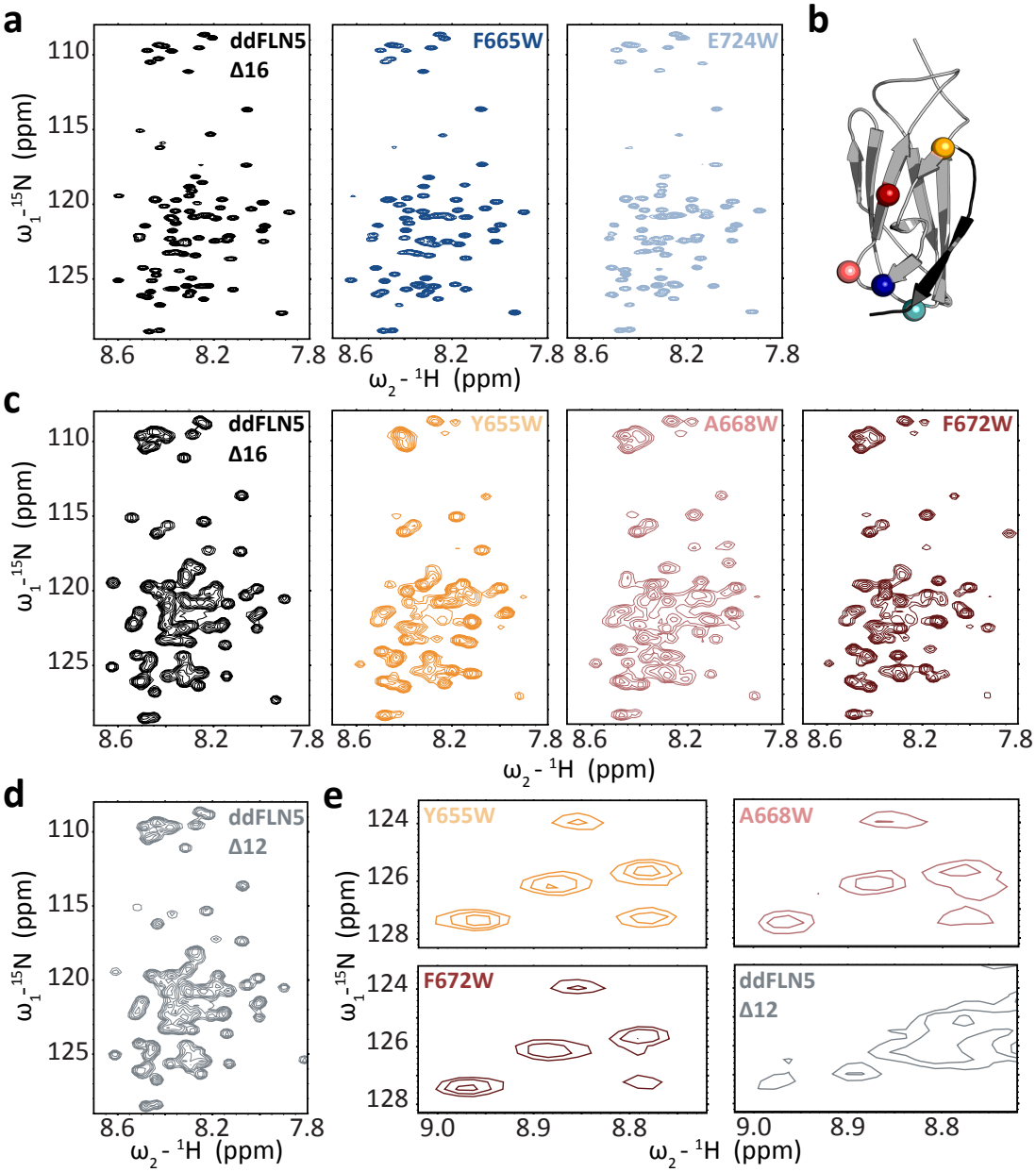

Figure 5.9: NMR characterisation of ddFLN5Δ16 tryptophan mutants a. ^{1}H-^{15}N HSQC spectra of WT ddFLN5Δ16 (black), F665W (dark blue) and E724W (light blue) at 10°C and 500 MHz.**b.** Structural distribution of mutation sites. Colours as in a, c and e. Black residues correspond to truncated residues in ddFLN5Δ16. **c.** ^{1}H-^{15}N SOFAST HMQC spectra of WT ddFLN5Δ16 (black), Y655W (orange), A668W (pink) and F672W (red) at 10°C and 700 MHz. **d.** ^{1}H-^{15}N SOFAST HMQC spectra of WT ddFLNΔ12 (grey) at 10°C and 700MHz. **e.** Cross-peaks arising from folded species in ^{1}H-^{15}N SOFAST HMQC spectra of WT ddFLN5Δ12 (grey), Y655W (orange), A668W (pink) and F672W (red) at 10°C and 700 MHz.

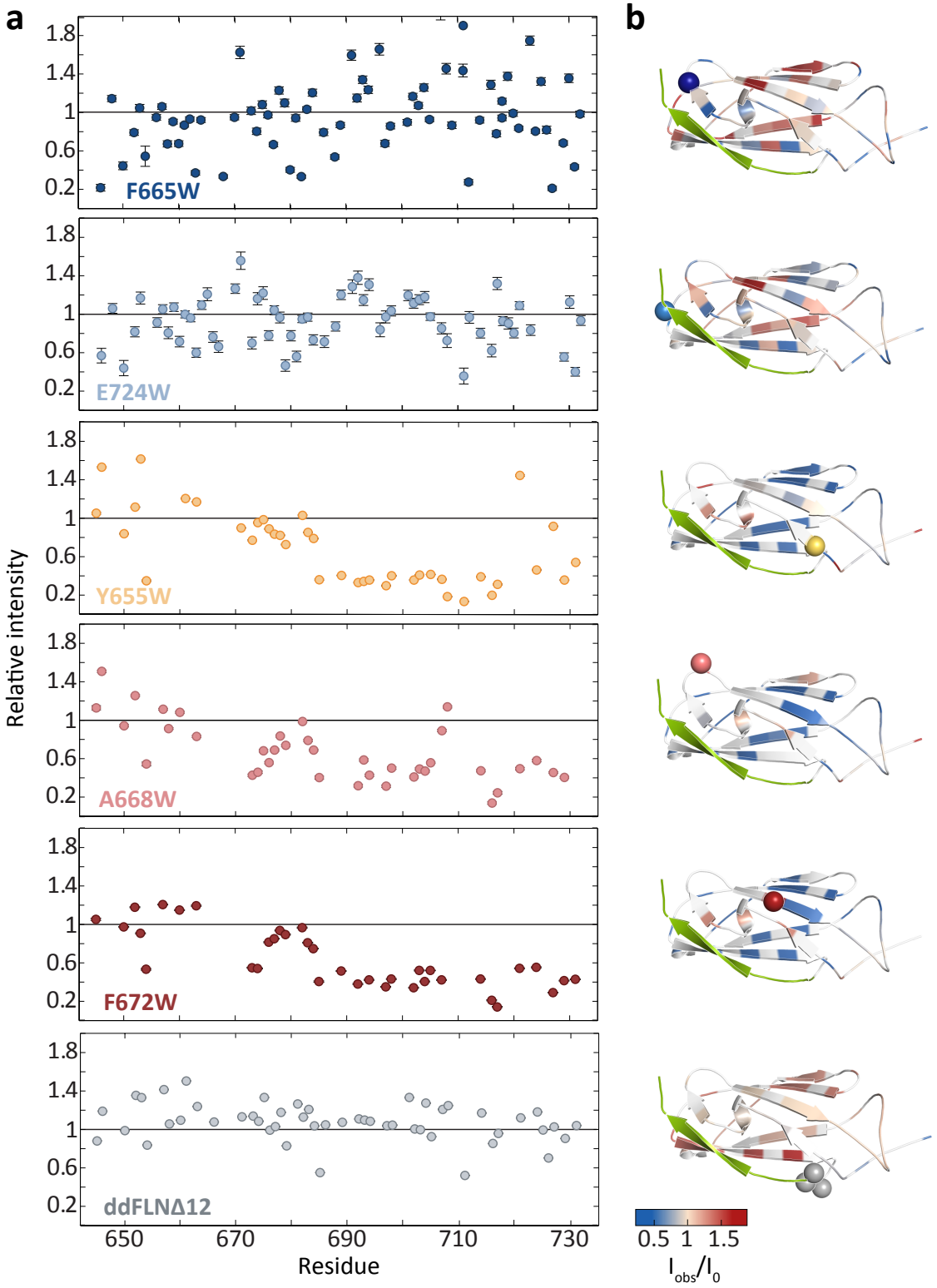

Figure 5.10: Resonance intensity analysis of ddFLN5Δ16 tryptophan mutants **a.** Relative resonance intensities of ddFLN5Δ16 tryptophan mutants relative to WT ddFLN5Δ16. **b.** Structural distribution of ddFLNΔ16 tryptophan mutants intensity variations in the full-length ddFLN5.

Further investigation is required to distinguish between the two hypotheses.

The observed disruptions to the NMR behaviour of ddFLN5 could be of both a structural and/or dynamic nature since NH chemical shifts are not sensitive reporters of secondary structure formation. The limited changes in the chemical shifts in the HSQC spectrs of the Trp mutsnts do however indicate that no significant tertiary structure has been formed in the disordered ddFLN5Δ16 construct by the insertion of a large aromatic residue at these sites. Further biophysical characterisation will be required to determine the nature of the observed perturbations and the effect on the overall stability of the full length ddFLN5 domain.

5.5.3 The interaction of TF with ddFLNΔ16 tryptophan mutants

^{1}H-^{15}N HSQC resonance intensities of ddFLN5Δ16 mutants F665W and E724W are unaffected upon equimolar addition of TF at 10°C (and 500 MHz). Addition of equimolar TF to Y655W, A668W and F672W however did result in notable changes in the ^{1}H-^{15}N SOFAST HMQC spectra. Selective broadening of residues was detected near the sites of mutation and in further regions of the protein sequence. Resonance broadening is strongest in the F672W mutant and clearly defines three sites affected by the presence of TF: residues 672-677 (broadened to 20%), the most hydrophobic region of the protein, residues 688-697 (broadened to 60%) and residues 714-719, the second site of above average hydrophobicity, broadened down to 30%. In Y655W and A668W, residues 672-677 are most affected by the presence of TF (signal intensities reduced to 50-60%) despite the sites of mutation not mapping to that exact region. The Y655W mutation appears to introduce a second more N-terminal site of TF interaction, around the site of mutation, between residues 655 and 668. A weak interaction with ddFLN5Δ12 (with 4 additional residue with respect to ddFLN5Δ) was detected at residues 672 to 677 with resonances broadened to 60%.

The absence of broadening in F665W and E724W indicates that introduction of tryptophans at these particular sites does not enhance the TF affinity for ddFLN5Δ16 despite the increased hydrophobicity, in particular for E724W. ddFLN5Δ16 mutants Y655W, A668W and F672W enhance the affinity of TF for the ddFLN5Δ16 sequence at residues 672 to 677 while further perturbations of resonance intensities appear mutant specific. The addition of 4 additional C-terminal residues appears to have a similar effect on the TF affinity for ddFLN5.

5.5.4 Discussion

The insertion of single tryptophan residues at specific sites (Y655, A668 and F672) of the ddFLN5Δ16 sequence introduced sites of TF interaction centred around a pre-existing

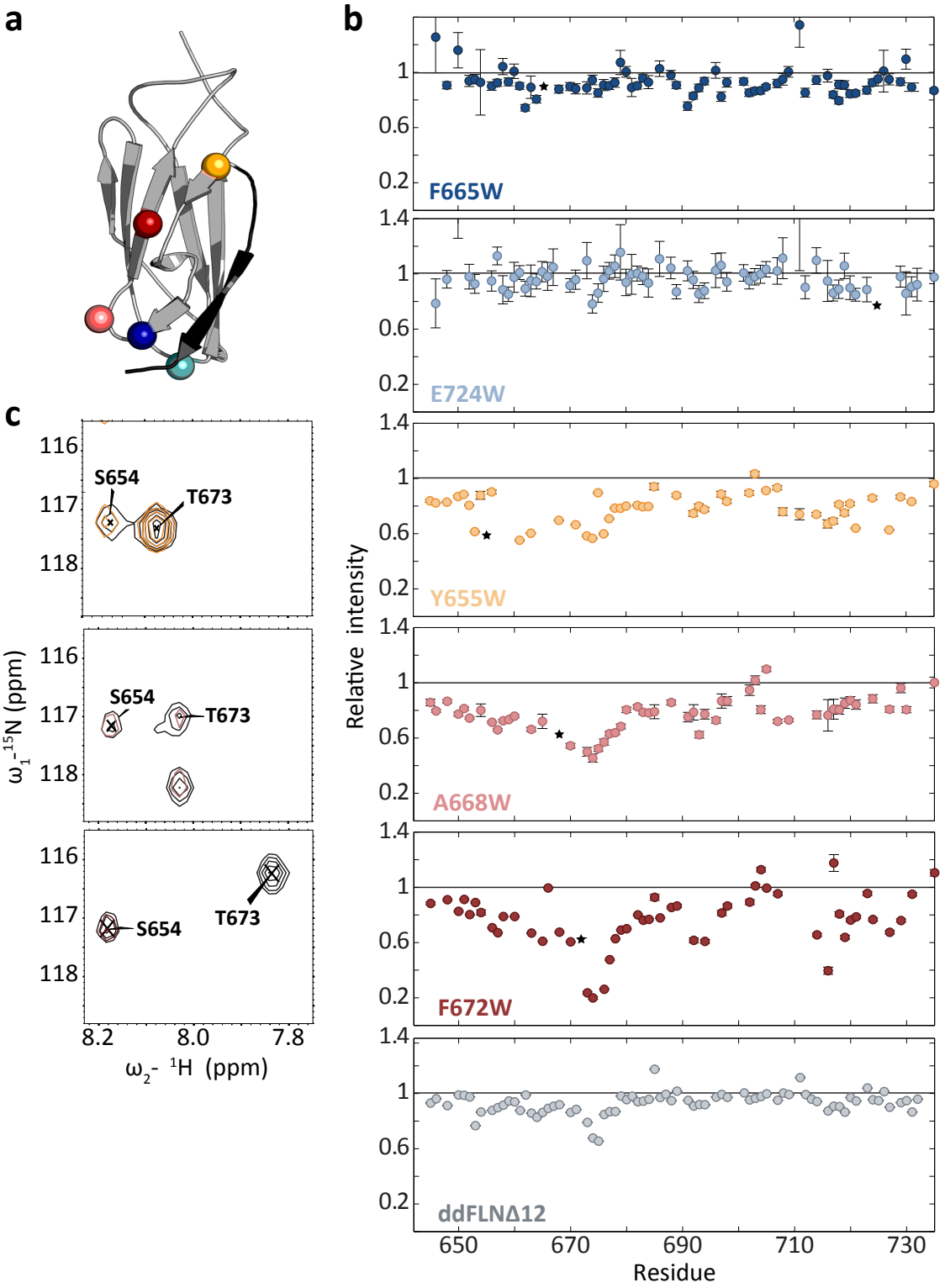

Figure 5.11: Resonance intensity analysis of ddFLN5Δ16 tryptophan mutants in the presence of TF. a. Relative resonance intensities of ddFLN5Δ16 tryptophan mutants in the presence of equimolar TF. **b.** Structural distribution of mutation sites. Colours as in a, c and e. Black residues correspond to truncated residues in ddFLNΔ16. **c.** Relative resonance intensities of ddFLN5Δ16 tryptophan mutant unassigned dispersed resonances in the presence of equimolar TF.

region of hydrophobicity as determined by amino acid analysis. A correlation between C-terminal exchange, the appearance of well-dispersed resonances and TF interaction was detected while no correlation with predicted increase in hydrophobicity was observed. The precise biological implications and the causal relationship between the current observations are unclear.

5.6 Co^{2+} paramagnetic relaxation enhancement measurements.

5.6.1 Introduction

Paramagnetic relaxation enhancement (PRE) measurements are a well established means of determining medium range distance constraints by NMR. The free electron of paramagnetic species such as a subset of transition metal ions induces rapid relaxation of the magnetism of nearby nuclei. The theoretical principles of the PRE effect are discussed in Section 2.1.6.5. PRE experiments have been extensively used in the study of residual structure in IDPs due to its capacity to detected short-lived conformational states.

Typical sources of paramagnetism used in NMR studies include native metal binding centres or nitroxide spin labels attached to engineered cysteine residues. In this study, all ddFLN5-derived TF substrate proteins contain N-terminal H_6-tags for affinity purification. Paramagnetic metal ions such as Cu^{2+}, Ni^{2+} and Co^{2+} readily coordinate to the indole rings of the H_6-tag (Fig 5.12).

In the study that follows, the possibility of investigating compaction in the ddFLN5+21 RNC and perturbations in its conformational ensemble in the presence of TF using metal ion PRE experiments is explored by characterising the interaction and PRE effect of metal ions binding to the N-terminal his-tag in the isolated ddFLN5 protein and truncation mutant ddFLN5Δ16.

5.6.2 Paramagnetic relaxation enhancement of Cu^{2+} ions in ddFLN5Δ16

Initially, the interaction of Copper(II) ions with ddFLN5Δ16 was investigated. The Cu^{2+} interaction with the H_6-tag is of high affinity but low specificity. This high affinity may be a favourable property in a PRE agent used in the study of RNC complexes since it is less likely to interfere with 70S and nascent chain stability.

Gradual titration of 100 μM ^{15}N ddFLN5Δ16 with 10, 20, 50, 75 and 100 μM $CuCl_2$ at 10°C and pH 7.5 resulted in extensive broadening of ^{1}H-^{15}N SOFAST HMQC resonances along the full ddFLN5Δ16 protein sequence (Fig 5.13). At concentrations below 50 μM $CuCl^2$ limited resonance broadening is observed to approximately 70% in the C-terminal half of ddFLN5Δ16 (685-734). Higher concentrations (50, 75 and 100 μM) result in the

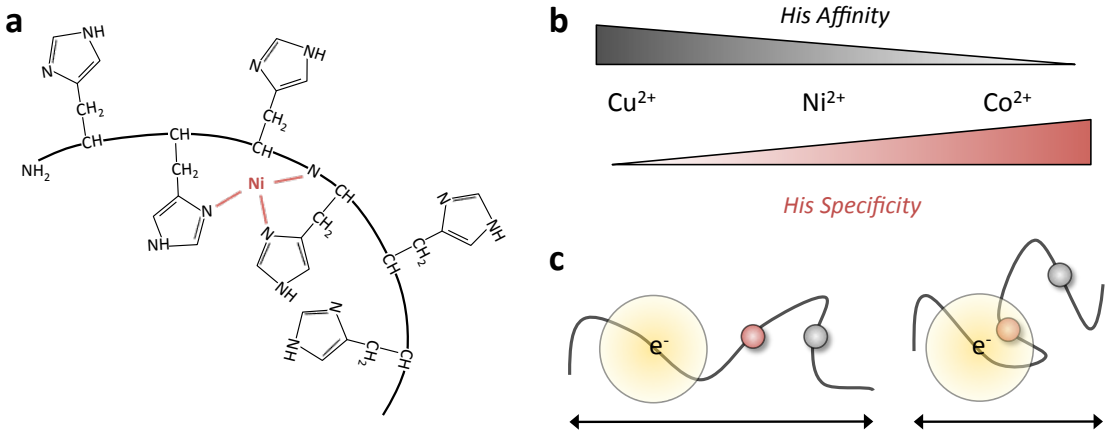

Figure 5.12: Paramagnetic relaxation enhancement through metal ion binding of the H_6-tag of ddFLN5Δ16. (**a**) Model of the coordination of Ni^{2+} by the H_6-tag. (**b**) The affinity of metal ions for histidine residues is inversely correlated to their specificity for histidine. (**c**) The PRE effect of the metal ions permits the detection of compaction in the ddFLN5Δ16 unfolded state.

broadening of resonances beyond detection at the ddFLN5Δ16 N-terminus with signals up to 40 % their original intensity detected for select resonances from residue 680 to 735. In particular at 1 equivalence, the highest concentration sampled, resonances appear to gain intensity progressively towards the C-terminus with intensities between 10-15%. Resonances originating from residues 724 till the C-terminus at 734 gradually recover to 25% (Fig 5.13.b).

The rapid transition in the observed resonance broadening with increasing Cu^{2+} concentrations (notably between 20 and 50 μM) suggests the presented data do not report on a single binding event of Cu^{2+} to the ddFLN5Δ16 H_6-tag only. This concentration dependent behaviour typically reflects cooperativity between multiple binding sites. The lack of defined structure in the ddFLN5Δ16 mutant does not readily translate into cooperative behaviour. Furthermore, the concentrations in question are substochiometric to the ddFLN5Δ16 present, indicating that the interactions are in fast exchange. It is therefore unlikely that the molecular mechanism would combine both weak, transient, interactions and cooperative behaviour. Another explanation for the rapid loss in resonance intensity is protein aggregation mediated by the interaction with Cu^{2+}. Addition of excess EDTA to 100 μM ddFLN5Δ16 and 100 μM Cu^{2+} however successfully recovered most ddFLN5Δ16 NMR resonances (residues 643-653 remain broadened beyond detection) indicating that the potentially observed aggregation is reversible.

The lack of characteristic PRE pattern in resonance broadening and unobservable resonances signify that Cu^{2+} ions are not suitable to the study of the ddFLN5+21 RNC

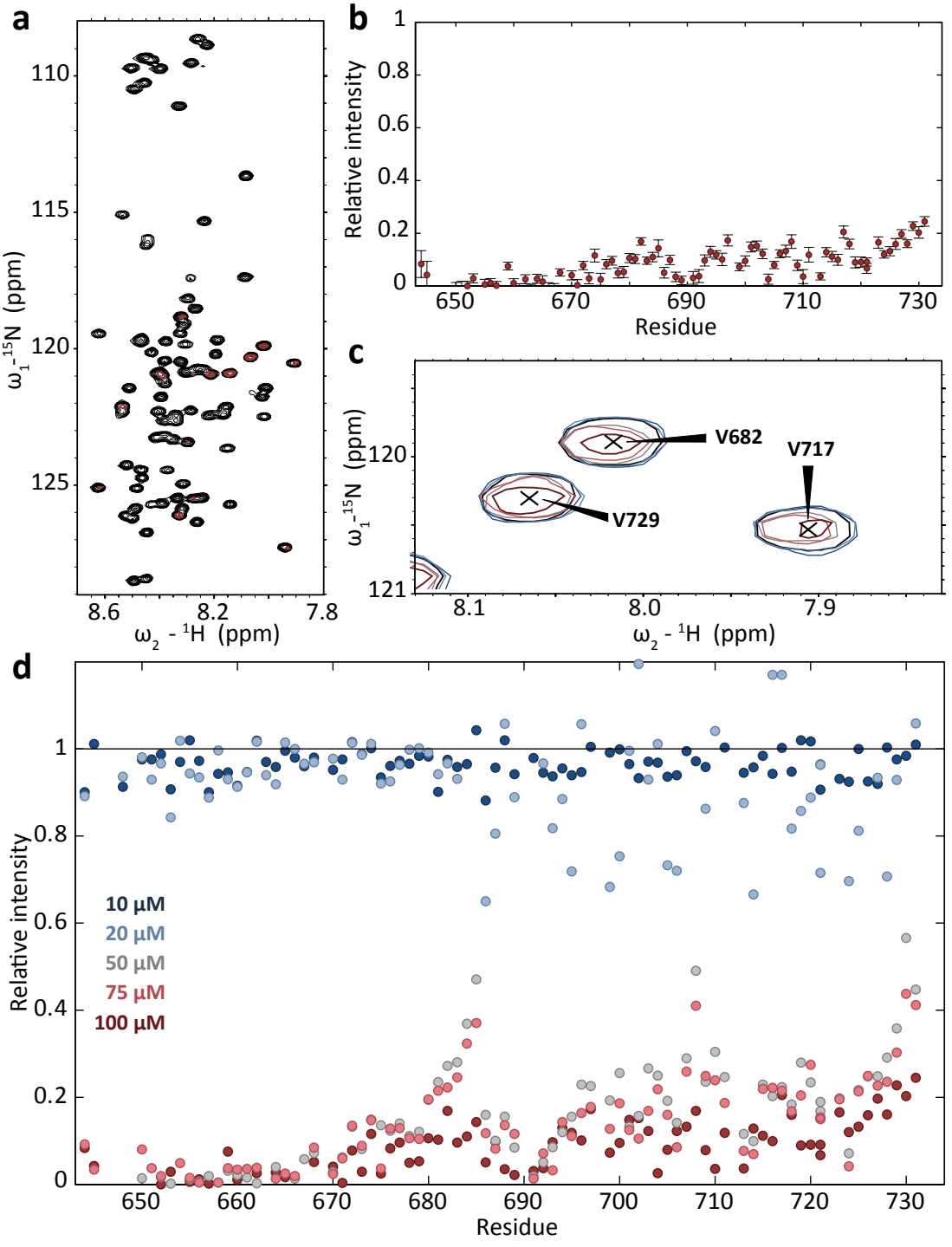

Figure 5.13: Titration of ddFLN5Δ16 with Cu^{2+} ions. a. Resonances in the ^{1}H-^{15}N SOFAST-HMQC spectra of 100μM ddFLN5Δ16 (black) are significantly broadened upon addition of one equivalence Cu^{2+} (red). **b.** Plot of the residual signal intensities in the presence of one equivalence Cu^{2+} shows all residues within ddFLN5Δ16 are affected by the presence of Cu^{2+} with signals slowly recovering towards the C-terminus of the protein. **d.** Plot of the residual signal intensities over the course of a gradual titration with Cu^{2+} shows a rapid disappearance of signal intensity at 0.5 equivalence.

by PRE.

5.6.3 Paramagnetic relaxation enhancement of Co^{2+} ions in ddFLN5Δ16

Cobalt(II) ions display weaker affinity for H_6-tags but greater specificity than Cu^{2+}. Titration of 100 μM ddFLN5Δ16 with Co^{2+} resulted in the selective broadening of resonances mapping to the N-terminus of the protein sequence, shown in Fig 5.14. The detected broadening gradually increases with increasing Co^{2+} concentration from the N-terminus where the H_6 tag resides. At 1 equivalence, broadening extents from the H_6-tag (638-643) (first detected residue is S645 at 40%) to residue T679. Additional broadening is detected at residue H653 which does not appear to propagate further into the sequence. Similarly, further C-terminal sites around residues D689 and D720 show additional broadening. Fit of the titrations data for residues 648 to two state exchange gives a dissociation constant of 15±5 μM for the interaction of Co^{2+} ions with the H_6-tag of ddFLN5Δ16.

The distance dependence observed in the broadening of signals in ddFLN5Δ16 upon interaction with Co^{2+} creates favourable conditions for the study of the unfolded state in the ddFLN5 nascent chain since the PRE effect of Co^{2+} ions can effectively be used as a molecular ruler between the N-terminus and any residues in the protein sequence. Additional sites of broadening around residues H653, D689 and D720 reflect either the presence of additional low affinity Co^{2+} binding sites or structural proximity to the ddFLN5Δ16 N-terminus. The affinity of residue 691 is calculated to be 25.5 ± 8.5 μM from the numerical fit of the titration data within error of the dissociation constant of the N-terminal interaction, suggesting that the broadening detected at these sites reflect the same binding event. Significantly more extensive broadening (beyond detection within 5 Å of contact site) would be expected for a direct interaction with Co^{2+} of this affinity.

In summary, the titration of 100 μM ddFLN5Δ16 with $CoCl_2$ results in the specific interaction of the Co^2+ ions with the H_6-tag of ddFLN5Δ16. Additional broadening observed in regions of the sequence distal to the N-terminus suggests compaction in the disordered ddFLN5Δ16.

In an attempt to simulate RNC conditions, the ddFLN5Δ16 concentrations were reduced to 10 μM and the titration repeated (See Fig 5.15). A similar N-terminal broadening pattern emerges over the course of the titration although the clear cut-off of the PRE effect at residue T679 does not emerge until excess Co^{2+} conditions are reached. At substoichiometric Co^{2+} concentration broadening does clearly propagate from the N-terminus but extends further into the sequence as more Co^{2+} is added. The numerical fit of this second data set produces similar values for the dissociation constant of the Co^{2+} binding reaction (8.6 ± 3.4 μM).

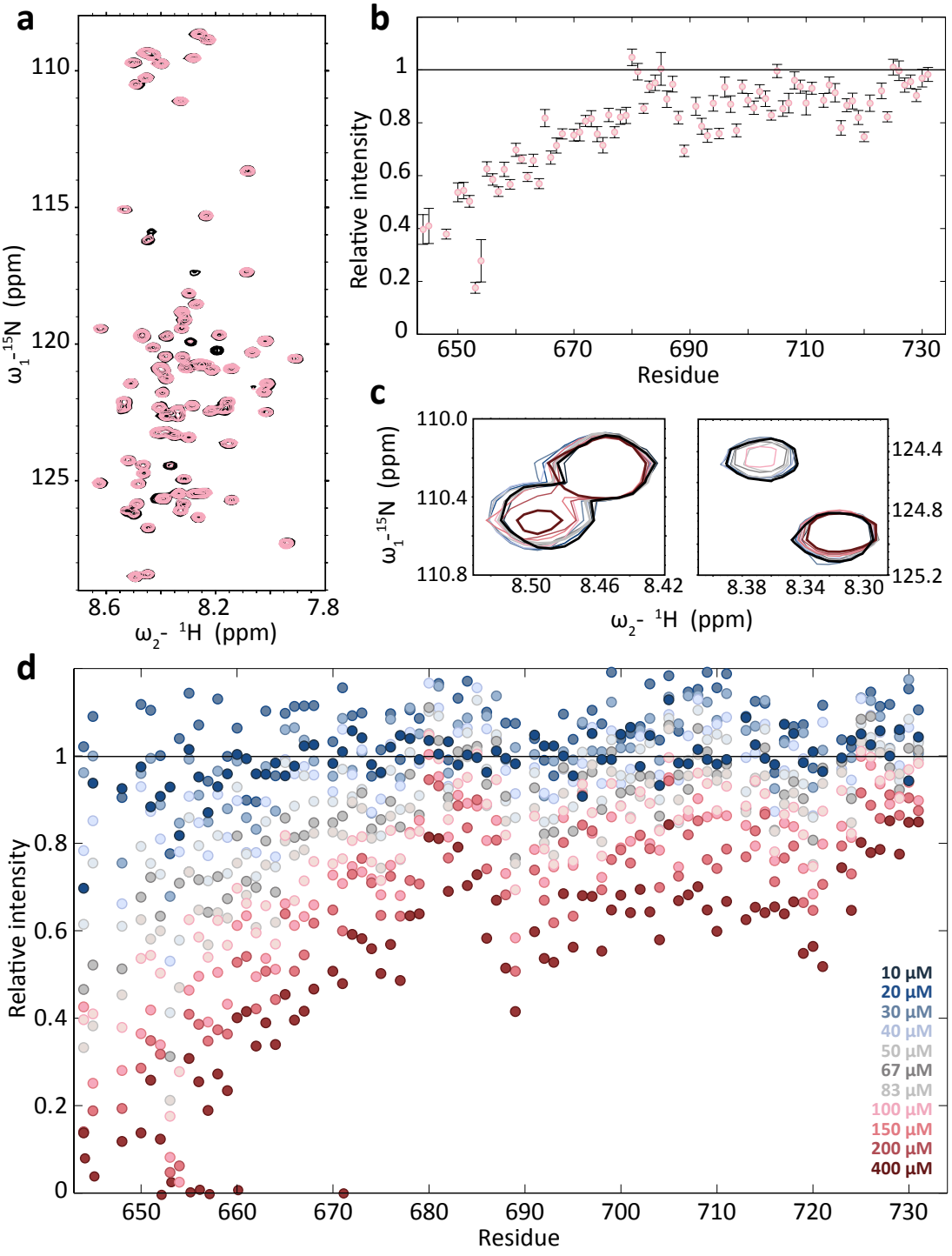

Figure 5.14: Titration of ddFLN5Δ16 with Co^{2+} ions. a. Select resonances in the ^{1}H-^{15}N SOFAST-HMQC spectra of 100μM ddFLN5Δ16 (black) are significantly broadened upon addition of one equivalence Co^{2+} (pink). **b.** Plot of the residual signal intensities in the presence of one equivalence Co^{2+}. Signals affected map to the N-terminus of the ddFLN5Δ16 sequence. **c.** Detail of ^{1}H-^{15}N SOFAST-HMQC resonances corresponding to residues G658, G725, K646 and F675 over the course of the Co^{2+} titration (colours as in d) **d.** Plot of the residual signal intensities over the course of a gradual titration with Co^{2+}.

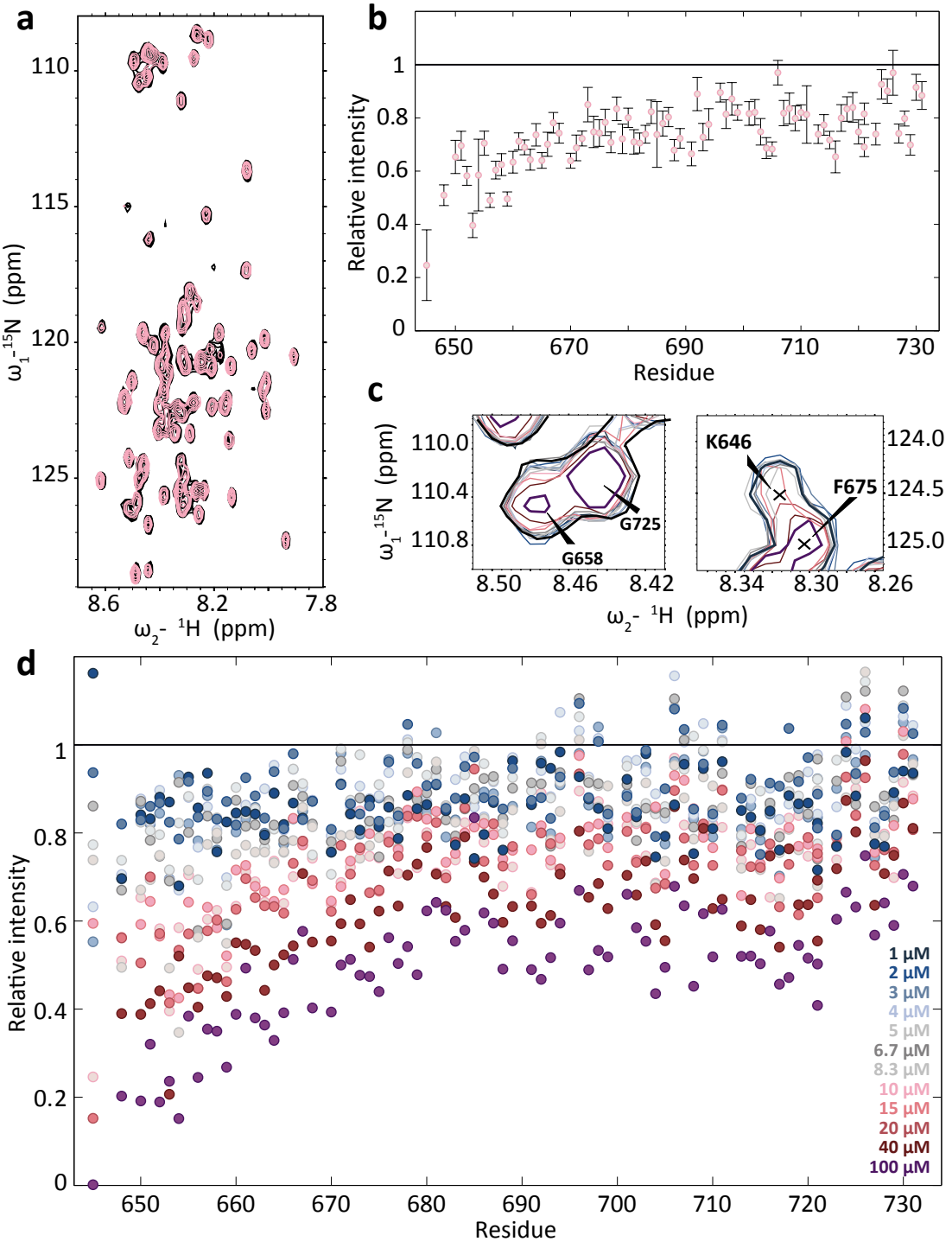

Figure 5.15: Titration of ddFLN5Δ16 with Co^{2+} ions. (**a**) Select resonances in the ^{1}H-^{15}N SOFAST-HMQC spectra of 10μM ddFLN5Δ16 (black) are significantly broadened upon addition of one equivalence Co^{2+} (pink). (**b**) Plot of the residual signal intensities in the presence of one equivalence Co^{2+}. Signals affected map to the N-terminus of the ddFLN5Δ16 sequence. (**c**) Detail of ^{1}H-^{15}N SOFAST-HMQC resonances corresponding to residues G658, G725, K646 and F675 over the course of the Co^{2+} titration (colours as in d) (**d**) Plot of the residual signal intensities over the course of a gradual titration with Co^{2+}.

5.6.4 Paramagnetic relaxation enhancement of Co^{2+} ions in ddFLN5

Full-length ddFLN5 adopts an immunoglobulin-like fold (Fig 1.8). Addition of Co^{2+} resulted in selective broadening of peaks in the ddFLN5 ^{1}H-^{15}N SOFAST-HMQC spectrum as shown in Fig 5.16.a. No clear sequence dependence of the broadening can be discerned from Fig 5.16.c. However, plotting the detected broadening on the ddFLN5 structure permits the identification of two distinct sites affected by the presence of Co^{2+}. Both the N-terminal and C-terminal hemispheres of the immunoglobin domain are affected by the presence of Co^{2+}. The effect of the N-terminal site can be readily explained by its close proximity to our predicted Co^{2+} interaction site, the H_6-tag. The detected broadening at the C-terminal hemisphere is however more difficult to explain. One possibility is a second binding site in this region of the protein.

Bioinformatic prediction of metal binding sites in the ddFLN5 structure identifies residues D666, E724 and E749 as having the correct orientation to form a metal binding triad. These residues are among those identified as affected by the presence of Co^{2+} indicating that they most likely do constitute a secondary metal binding site, in competition with the H_6-tag in the folded ddFLN5.

Affected resonances moreover display chemical shift changes over the course of the titration with Co^{2+} (Fig 5.16 **b**). Unclear is whether these chemical shift changes report on the fast timescale of the ddFLN5 interaction with Co^{2+} or wether the detected chemical shift perturbations correspond to pseudo-contact shifts.

5.6.4.1 Discussion

The folded ddFLN5 population can be readily distinguished from the unfolded ddFLN5Δ16 state by the observed broadening pattern upon addition of Co^{2+}. In the disordered ddFLN5Δ16, Co^{2+} ions specifically interact with the H_6-tag with an affinity of $\sim$10 μM. The broadening pattern is conserved across concentrations of ddFLN5Δ16 once saturation is achieved. The Co^{2+} paramagnetic relaxation broadening pattern thus appear to to serve as a molecular ruler for the compaction of the unfolded species of ddFLN5.

Coordinated Co^{2+} is highly anisotropic and large pseudo-contact shifts may be expected. No chemical shifts perturbations were however detected for the disordered ddFLNΔ16 indicating the protein tumbles independently from the H_6-tag and no preferred conformation (of the H_6-tag) is detected.

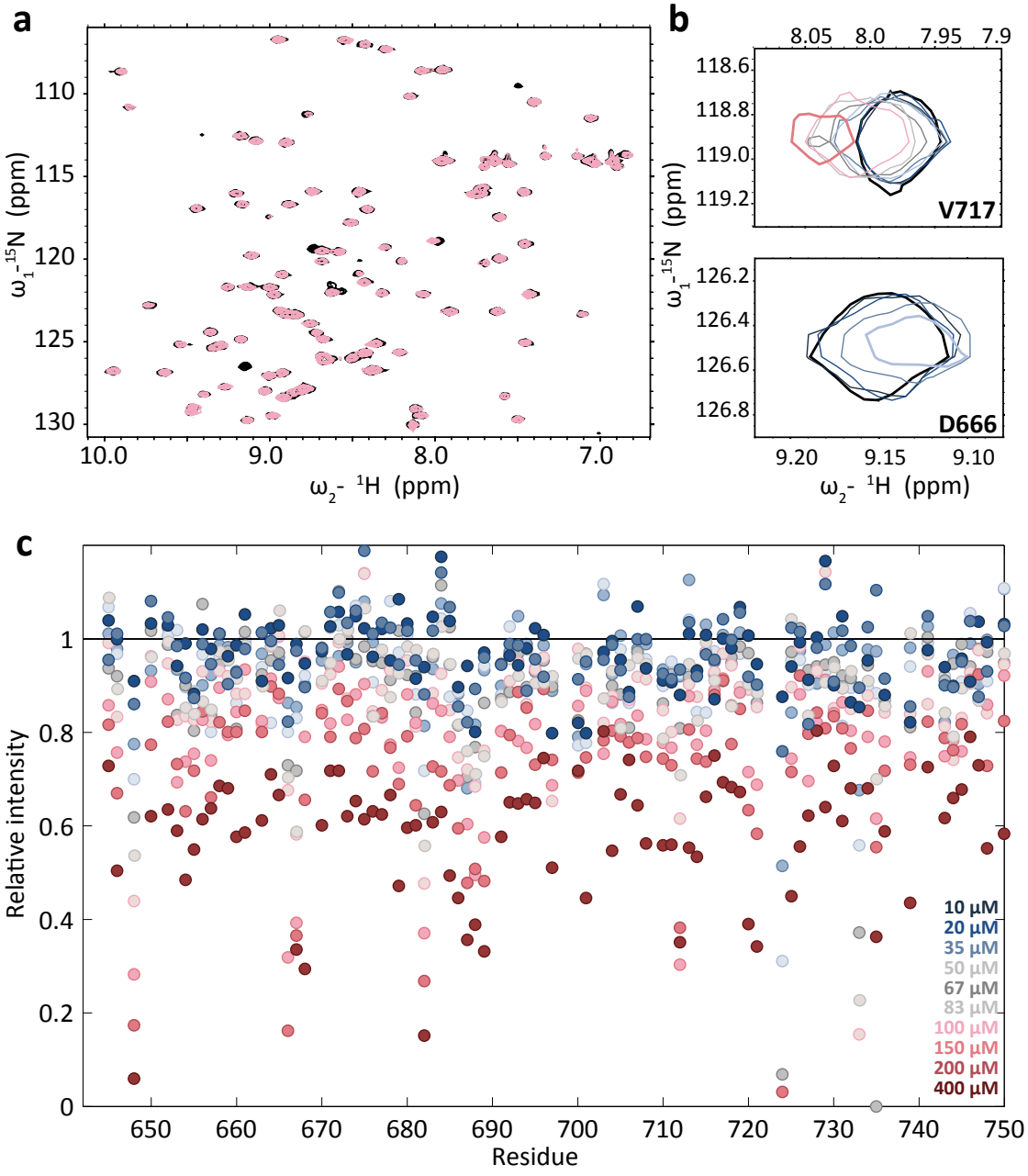

Figure 5.16: Titration of ddFLN5 with Co^{2+} ions. a. Select resonances in the ^{1}H-^{15}N SOFAST-HMQC spectra of 100μM ddFLN5 (black) are significantly broadened upon addition of one equivalence Co^{2+} (pink). **b.** Detail of ^{1}H-^{15}N SOFAST-HMQC resonances corresponding to residues D666 and V717 over the course of the Co^{2+} titration (colours as in c) **c.** Plot of the residual signal intensities over the course of a gradual titration with Co^{2+} shows selective broadening of resonances originating from non-neighbouring residues.

5.6.5 ddFLN5Δ16 and silent 70S ribosomes titration with Co^{2+}

The titration of isolated ddFLN5Δ16 provides encouraging results in terms of the application of the Co^{2+} PREs from the ddFLN5 H_6-tag as a measure of compaction in disordered nascent chains. The 70S ribosome however, interacts extensively with metal ions, in particular Mg^{2+}. These ribosome metal-binding sites might therefore compete with the H_6-tag for Co^{2+}.

Practically, 10 μM 70S were added to 25 μM ^{15}N ddFLN5Δ16 at pH 7.5 and 10°C. Co^{2+} was gradually added up to 40 fold excess over a 6 point titration (Fig 5.17). Again, selective broadening was detected at the N-terminus of the ddFLN5Δ16 sequence. However, much higher concentrations of Co^{2+} were required to attain the same extent of broadening. Indeed, until one molar equivalence, ddFLN5Δ16 resonances are not affected by the addition of Co^{2+}. At higher concentrations Co^{2+} (4x and 10X) broadening is detected at the N-terminus affecting residues up to A648. At 40 fold excess the broadening pattern of ddFLN5Δ16 is very similar to that in the absence of 70S extending to residue T679 in a distance dependent manner (Fig 5.17).

The presence of ribosomes thus competes with the interaction of Co^{2+} and the N-terminus of ddFLN5Δ16 but does not affect the nature of the interaction. Moreover, this study found that ddFLN5Δ16 does not interact with the 70S ribosome as can be observed from the first titration point, and the ribosome does not affect the degree compaction in the N-terminal region of the unfolded ddFLN5Δ16 indicating the study of Co^{2+} PREs in the ddFLN5+21 RNC may be feasible.

5.7 Concluding remarks

The study of various ddFLN5 constructs in the presence of TF showed the ddFLN5 sequence does not carry any native TF interaction sites. Interaction could be selectively introduced by single amino acid substitutions to tryptophan at select sites around regions in the ddFLN5 sequence of above average hydrophobicity. The interaction of TF with the ddFLN5Δ16 tryptophan mutants was selective for residues 671-676 with limited propagation of the resonance broadening along the sequence. Only three out of 5 produced Trp mutant interact with TF: 655W, A668W and F672W. Unclear is the driving force behind the interaction since no correlation to increased hydrophobicity, sequence amino acid composition or effect on the stability of the folded ddFLN5 domain was observed.

The specific interaction of Co^{2+} with the ddFLN5Δ16 H_6-tag provides an interesting avenue for the investigation of compaction in the disordered protein and its nascent chain equivalent ddFLN5+21 RNC. This preliminary study detected a PRE up to 40 residues removed from the H_6-tag suggesting significant compaction of the ddFLN5Δ16 N-terminus

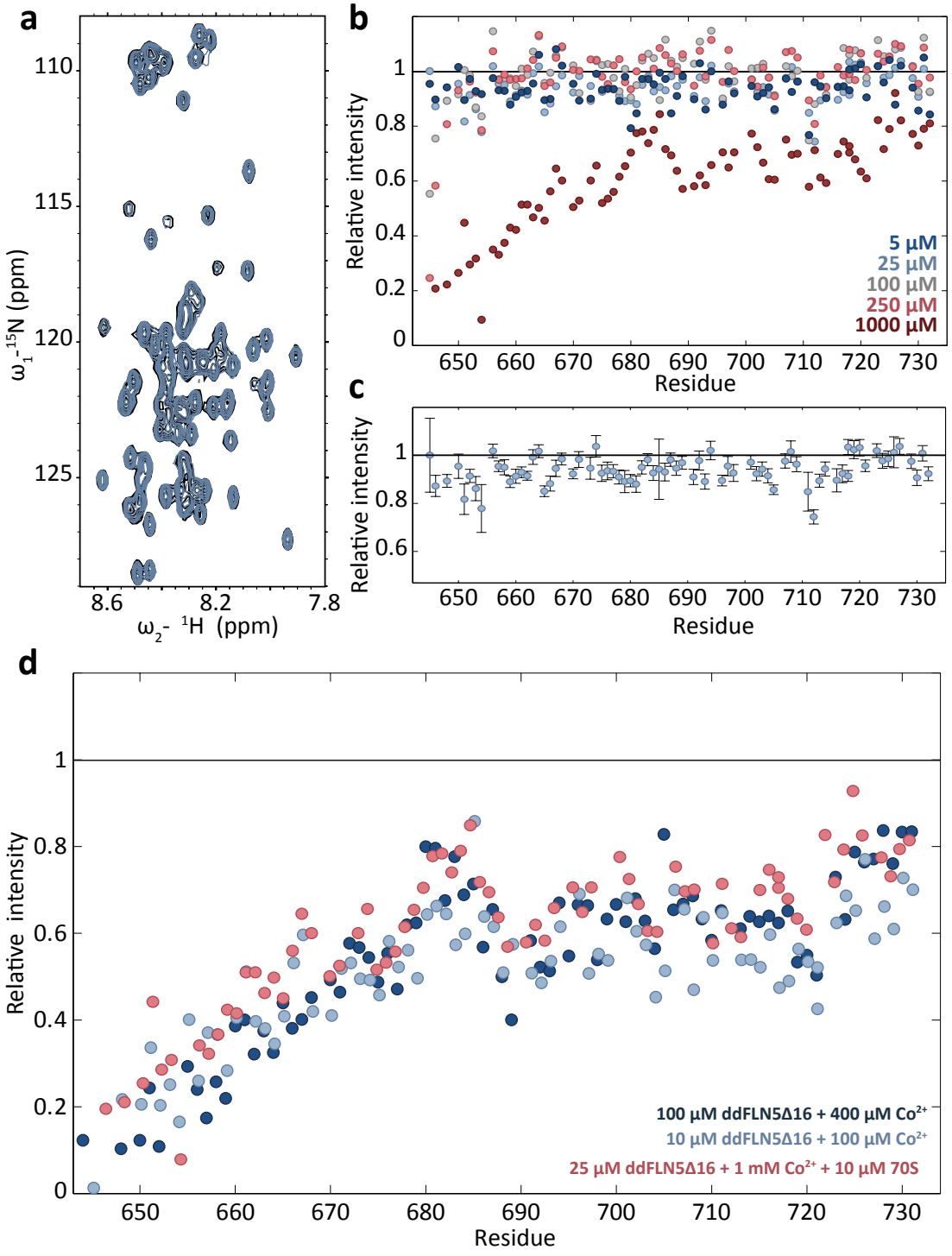

Figure 5.17: Titration of ddFLN5Δ16 with Co^{2+} ions in the presence of ribosomes. (**a**) Resonances in the ^{1}H-^{15}N SOFAST-HMQC spectra of 25μM ddFLN5Δ16 in the presence of 10μM 70S (black) are not significantly perturbed upon addition of one equivalence Co^{2+} (light blue). (**b**) Plot of the residual signal intensities upon progressive addition of Co^{2+}. (**c**) Plot of the residual signal intensities in the presence of one equivalence Co^{2+}. (**d**) Plot of the residual signal intensities over the course of a gradual titration with Co^{2+}.

. Secondary sites of broadening detected at D689 and D720 indicate that compaction might not be limited to the N-terminus only since the calculated dissociation constant at these sites is in good agreement with values found at the N-terminus $\sim$15μM. C-terminal broadening to 60% was observed for the above mutants relative to the wild-type ddFLN5Δ16 indicating exchange occurs. Co^{2+} PREs may be able to define the observed exchange, since changes in the compactness of the disordered state should give rise to a different broadening pattern.

The ddFLN5+21 RNC was shown to produce high quality spectra at 10°C. Even in the presence of TF, where signals are broadened to 65%, all resonances remain observable. This suggest it may be an ideal system to study the compaction of the nascent chain by Co^{2+} PREs because a considerable amount of signal intensity is required to detected residues specific broadening in particular since the effect of TF in the compaction is likely to be small in the absence of specific interaction. The use of Co^{2+} PREs as molecular ruler for compaction in disordered states thus opens up number of avenues for further investigation.

Chapter 6

Discussion

This thesis presents a strategy for the residue-specific investigation of the dynamic interaction of TF with ribosome nascent chain complexes and advances our understanding of the role of this chaperone in the process of co-translational folding. A brief summary of the results presented in this thesis is provided here, followed by a more in depth discussion of individual findings.

In Chapter 3, an experimental protocol was developed for the selective protonation of side-chain methyl groups in perdeuterated TF, allowing the exploitation of the methyl TROSY effect to improve the quality of the NMR spectrum of this large 96 kDa dimeric macromolecule. This sensitivity enhancement permitted a detailed study of the thermodynamics of the TF dimerisation reaction, the proposal of a kinetic model for the dynamic self-association of the TF, and a solution structure model of the TF dimer.

Chapter 4 and 5 reported the investigation of the TF interaction with two model nascent chain substrates: α-synuclein and ddFLN5. We found that the presence of TF significantly perturbs these ribosome-associated nascent chain substrates even in the absence of predicted TF interaction sites. The study of the analogous isolated proteins revealed that only limited interaction occur in the absence of the ribosome but that TF recruitment could be induced by the introduction of aromatic residues at key positions of the α-synuclein and ddFLN5 sequence.

6.1 TF structure and dynamics

6.1.1 Selective protonation of methyl groups permits the detailed characterisation of TF by NMR spectroscopy

High quality NMR spectra could be obtained from the 48 kDa macromolecule, TF, using selective labelling strategies pioneered in the last decade (Fig 3.5, Section 3.3). The selective protonation of long side-chain methyl groups on a perdeuterated background allowed improvements to the NMR sensitivity of ca. 13 fold for TF^{ile} relative to uniform ^{13}C labelled TF (Section 3.3.3) and the usage of TROSY optimised pulse sequences enabled a highly detailed NMR characterisation of the dynamics of self-association (Section 3.5.3), the thermodynamics of the monomer/dimer equilibrium (Section 3.5.2) and the dimer solution structure (Section 3.5.1).

Although TF is stable on the relative timescale of the NMR experiments with a lifetime of 50 hours at 25°C, TF is rather unstable relative to typical proteins investigated by NMR. It is likely that the multi-domain and extended organisation of the TF structure results in a protein more prone to degradation by proteolysis. In the context of the study of the TF interaction with nascent chains, the stability of the RNC is however limiting at 9-18 hours at 10°C for the ddFLN5+21 RNC investigated in Section 5.4.

6.1.2 Mathematical model of the complex TF equilibrium

NMR spectroscopy provides insights into structural and dynamic properties of an ensemble average of states. In the case of TF, the observed signals thus contain information regarding all the TF states present. This constitutes a difficulty in attributing specific observations to one particular exchange event such as for instance the simultaneous monomerisation and substrate interaction in the presence of α-synuclein (Luc 87-100), described in Section 4.4.3. A highly detailed understanding of each of the underlaying reactions is therefore required to deconvolute the various contributions to the NMR signal.

The mathematical model presented in Section 3.4 permits the prediction of changes in TF population under specific experimental conditions based on currently available thermodynamic information. In particular in the work on the TF-ribosome interaction (Section 4.5.1) the reconstruction of predicted NMR signals based on population distributions proofed essential. Moreover, it provides an insight into the changes to TF population distribution under different scenarios within the cell, highlighting the sophistication of the interplay between states, discussed further Section 6.3.

6.1.3 Solution structure of TF dimer

A combination of NMR methods (Section 3.5.1) that included intensity analysis, cross-saturation and paramagnetic relaxation enhancement (PREs), enabled the determination of the interface of TF dimerisation and the proposal of a specific orientation of the two TF molecules within the dimeric complex in solution. Residues involved in dimerisation as identified by cross-saturation (Section 3.5.1.2), map to all three TF domains (Fig 3.16), specifically the PPIase active site, the arm 2 region of the SBD and the inward facing RBD surface, shown in yellow Fig 6.1.a.

Intensity analysis identified a more extensive TF surface involved in the dimerisation process (Section 3.5.1.1 and Fig 3.15), with particular strong broadening of both RBD and SBD inner surface resonances, which is in good agreement with chemical shift perturbations measured by Saio *et al.* upon TF monomerisation [130]. While intensity analysis reports on all conformational changes upon TF dimerisation, cross-saturation is limited to interface resonances only. The additional sites identified by intensity analysis thus report on changes within the relative arrangement of TF domains upon dimerisation.

Inter-domain flexibility has previously been proposed based on the comparison of various TF monomer crystal structures shown in Fig 3.22 [64, 84, 126] and the present study has allowed us to understand these conformational changes at residue-specific resolution. Indeed, the additional sites of resonance broadening reported by the resonance intensity analysis described here (Section 3.5.1.1) map to hinge regions between the RBD and SBD, the two arm regions of the RBD and the SBD to PPIase interface ('neck' region).

Medium range (5-20 Å) distance measurements obtained by PRE experiments of PROXYL spin-labelled TF mutants (Section 3.5.1.3), restrains the relative orientation of the two TF molecules within the dimer. A model of the proposed solution structure of dimeric TF is discussed in Section 3.5.1.4 and reprised in Fig 6.1.a. This preliminary model proposes a head-to-toe symmetrical arrangement with principal contacts formed by the extremities of the RBD and SBD 'arm' regions and secondary contacts formed by the PPIase domain. The proposed conformations resembles that of the tetrameric TF_2:$S7_2$ crystal structure, Fig 6.1.b [84]. The crystal structure does not, however, satisfy all experimental constraints, in particular the PREs detected in the PPIase domain from all 4 spin label positions (14, 150, 326 and 376, Section 3.5.1.3) which places this domain in a more central position than it occupies in the S7-bound crystal structure. Within the tetrameric complex, the movement of the PPIase domain is obstructed by the presence of the S7 substrate. It is therefore likely that the absence of S7 permits a tighter packing of the two TF molecules, placing the PPIase domain near the RBD-SBD hinge region of the second monomer as predicted by the presented structural model (Fig 6.1.a).

Restrained computational docking of the two TF molecules in HADDOCK [244, 245]

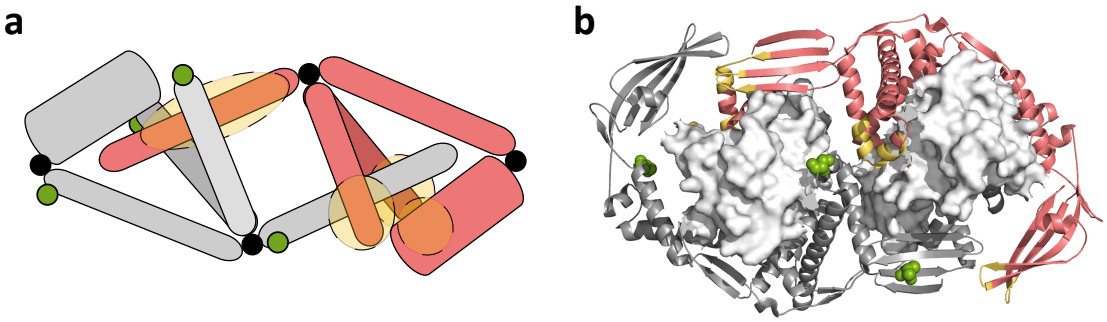

Figure 6.1: Structural comparison of proposed TF dimer structure and S7-bound TF. (**a**) Proposed model for the solution structure of dimeric TF based on NMR data presented in this study. Black dots correspond hinge regions linking the various TF domains while green dots correspond to sites of proxyl-labelling for PRE experiments. Yellow circles define identified dimerisation interfaces. (**b**) Crystal structure of *T. maritima* TF in complex with ribosomal protein S7 (pdb 3GTY) [84]. Colouring as in a.

based on the experimental data described in this thesis is currently in progress and is likely to improve the structural model of the TF dimer in solution to a near residue-specific level.

6.1.4 Conformational change within the TF dimer

The extensive broadening of RBD resonances (Section 3.5.1.1) and the absence of clear distance distribution detected within the TF dimer by EPR experiments (Section 4.4.4.1) indicate that TF may not adopt a rigid dimeric structure in solution but instead constitutes a dynamic ensemble of conformations around the structural model proposed here (Fig 6.1.a). Indeed, while both cross-saturation (Section 3.5.2.1) and CPMG RD (Section 3.5.3.2) experiments strongly implicated all TF domains including the PPIase, in the dimerisation interface, PPIase resonances give rise to expected signal intensities over the course of a dilution series (Section 3.5.1.1). On average, the chemical environments of PPIase residues are thus not strongly perturbed upon dimerisation. In order to reconcile both observations, we propose that the RBD and SDB constitute the principle dimeric interface while the PPIase forms transient interactions and is structurally dynamic. This is in agreement with previous observations across multiple crystal structures that the PPIase domain can rotate up to 25° around its flexible linker (Fig 3.23). An interesting feature of the proposed conformational freedom of the PPIase domain is that it can still function both as substrate binding site and catalytic PPIase within the TF dimer. This suggests a more extensive role for the excess TF in the cytosol and a potential TF PPIase activity independent from interaction with the ribosome.

Hinge regions between domains and linking the two arms of the SBD were further shown to undergo conformational change upon dimerisation, without being involved in

the direct dimer interface. More wide-spread conformational change are thus likely to occur in the transition from monomeric TF to the dimer.

6.1.5 Study of TF dynamics reveals step-wise mechanism of dimerisation

The broadening of RBD resonances upon TF dilution at 900 MHz, observed by intensity analysis (Section 3.5.1.1) and highlighted in the study of the TF interaction with PhoA [130], indicate the TF dimer is a highly dynamic system. The presence of millisecond dynamics was therefore investigated by CPMG relaxation dispersion measurements in order to investigate the apparent exchange processes in more detail.

CPMG relaxation dispersion measurement at multiple field strengths and temperatures (Section 3.5.3.2) found that millisecond dynamics were widespread, mapping to all TF domains. Exchange contributions to the transverse relaxation rate (R_{ex}) were particularly large in the RBD and arm 2 regions of the SBD but further dispersions were found in the SBD cavity, arm 1 and neck regions, as well as the PPIase domain (Fig 3.24). The sites of millisecond dynamics identified in Section 3.5.3.2 correspond to TF regions involved in dimerisation as determined by intensity analysis (Section 3.5.1.1). This suggested that the observed exchange processes were associated with the dimerisation reaction. However, the observed millisecond timescale dynamics were however found to be concentration independent (Section 3.5.3.2) indicating that the observed dispersion do not report directly on the dissociation of the two TF monomers. Indeed, the half-life of the TF dimer is approximately 1 s^{-1} [148], is approximately two orders of magnitude slower than exchange rates of 400 s^{-1} determined by relaxation dispersion measurements.

We therefore hypothesised that the observed millisecond dynamics in TF report on the independent association and dissociation of two separate symmetric dimerisation interfaces (Fig 3.25). This is in agreement with the structural model of the TF dimer proposed in Section 3.5.1.4, where two symmetrical contacts form the main dimerisation interface. The excited state was found to be populated to 6% at each of the two dimerisation interfaces and to exchange with the major state with an exchange rate of 425 s^{-1} (Section 3.5.3.2). The experimental dissociation rates of each interface is therefore 25 s^{-1} giving rise to a global dissociation rate of the TF dimer of 1.5 s^{-1}, in good agreement with the half-life determined using fluorescence anisotropy [148]. We therefore propose a model for the step-wise association and dissociation of the TF dimer (shown in Fig 3.25.d) based on the formation and dissociation of two independent but symmetrical dimerisation interfaces. Only the simultaneous dissociation of both sites results in complete TF monomerisation.

6.1.6 Quantification of concentration dependent changes by whole-spectrum analysis

The dissociation constant of TF dimerisation was determined independently by linear regression analysis of HMQC signal intensities and line-shape analysis (Section 3.5.1.2). The obtained values of 7.2 and 10.5 μM respectively are greater than those described in previous studies of the dimerisation reaction summarised in Table 3.1 [148, 159, 160]. The likely origin of the discrepancies is the use of D_2O through-out this study weakens hydrophobic contacts relative to H_2O via the solvent isotope effect on H-bond strength.

The linear regression analysis developed in this thesis has the unique advantage to monitor concentration dependent changes across the spectrum. Typical quantification of NMR titration data focusses on a number of well resolved resonances, and in the slow exchange regime often monitors one state only across the titration. Due to extensive chemical exchange broadening of the TF dimer resonances and small chemical shift differences between monomeric and dimeric TF, identification of multiple well-resolved resonance pairs proved difficult.

Our novel method evaluates the concentration dependent changes to the NMR signal at each coordinate of the HMQC spectrum, thus including low signal to noise data from the minor species permitting more accurate determination of the dissociation constant of the reaction. It does not require the attribution of resonances to either of the two states nor accurate determination of resonance maximum intensities or signal integrals. Moreover, potential biases introduced by chemical exchange broadening or unobservability of the minor state are by-passed in this whole-spectrum analysis. We therefore believe that the presented method for quantification of NMR titration data will be highly applicable to a diversity of systems currently under investigation by NMR

6.1.7 Deconvolution of monomer and dimer contributions to the observed NMR signal

The linear regression analysis of the entire NMR spectra (Section 3.5.1.3) as opposed to the more typical approach of evaluating individual resonance signal changes across a titration permitted the deconvolution of the contribution of the monomeric and dimeric TF based on their behaviour across concentrations (Fig 3.18). This is to our knowledge the first example where such an approach has permitted the determination of the NMR spectra of two exchanging species where one state can not be isolated under condition suitable to NMR. Here, the TF monomer is unaccessible at the relative high concentrations required for NMR measurements as described by the mathematical model developed in Section 3.4.2. While the NMR study of the TF interaction with PhoA [130] overcame this problem by investigating a truncated TF product, the spectra presented here correspond

to full-length WT TF in the absence of substrates.

The quantification method as applied in this thesis (Section 3.5.1.3) is however solely applicable to systems showing slow exchange behaviour. While the majority of the observed resonances in the TF^{ILV} HMQC spectrum at 900MHz are indeed in the slow exchange regime between monomeric and dimeric TF states, the study of the TF interaction with PhoA [130] found significant exchange broadening of RBD resonances in the TF dimer. Since these resonances are broadened beyond detection under our experimental conditions, their concentration dependent behaviour could not be analysed using the described linear regression approach. These resonances do therefore not appear in the deconvoluted monomer and dimer spectra presented in Section 3.5.1.3.

6.1.8 Competition of TF dimerisation with other substrate interaction equilibria

The model of the solution structure of the TF dimer proposed in Section 3.5.1.4 explains previous observations made about the competition of the various TF interaction equilibria. First of all, the proposed dimer structure (Fig 6.1.a) occludes the characteristic TF ribosome binding motif through the conformational change of the PPIase domain. The TF interaction with the ribosome thus would occur with the monomeric state as has been previously reported [159, 160].

Furthermore, the interaction of high affinity substrates (e.g. α-syn (Luc 87-100), Section 4.4.2.3) with TF was found to induce monomerisation. The solution structure model of the TF dimer indicates that substrate interaction sites as identified by Saio *et al.* [130] are buried within the interface. Isolated protein substrates thus compete with the dimerisation reaction for interaction at specific sites. The proposed step-wise dissociation of the dimer outlined in Section 3.5.1.4 further explains the observation that the interaction with weak protein substrates (H_6-α-synuclein) occurs with dimeric TF (Section 4.4.1). The proposed pathway for TF dimerisation therefore introduces a mechanism for competition with other substrate interactions distinct from the typical view of monomer interaction with substrate.

It appears that the dynamic nature of this particular TF reaction is highly conducive towards its function as promiscuous co-translational chaperone. The dynamic nature of the dimerisation and isolated protein interactions means that cellular pools of TF are never trapped in a particular substrate-bound state. The constant competition between various substrates imply a high degree of adaptability of TF and rapid response to sudden changes in the cellular environment.

6.2 TF interaction with nascent chain substrates

6.2.1 Choice of TF substrates

The heterologous substrate systems investigated in this study, α-synuclein and ddFLN5, were chosen based on a pre-existing understanding of their behaviour by NMR both as isolated proteins and as nascent polypeptides, providing a reference frame for the interpretation of their behaviour in the presence of TF [65, 258] (Waudby *et al.*, in preparation, Cabrita *et al.*, in preparation). Despite both proteins' eukaryotic origins, the promiscuity of TF observed in the bacterial cell suggests that any mechanistic insights developed from the study of the interaction of TF with these substrates can indeed be generalised. Eukaryotic and other non-native substrates to the *E.coli* TF have previously been used successfully to investigate length and sequence dependence of the TF interaction with substrates and its effect on the folding state of nascent polypeptide chains on the ribosome [144].

6.2.2 TF preferentially interacts with sequences enriched in aromatic residues

The TF interaction with α-synuclein is of a highly transient nature as characterised by line-broadening in the α-synuclein ^{15}N HSQC spectrum (Fig 4.2, Section 4.4.1). A specific but low affinity interaction was identified at the α-syn N-terminus centred at Phe 4 (Section 4.4.1). The substitution of a highly hydrophobic sequence from firefly luciferase dramatically enhanced the TF affinity for α-synuclein (Section 4.4.2.3) and in a similar fashion, TF could be recruited to isolated ddFLN5δ16 by the insertion of tryptophan residues at key hydrophobic sites (Section 5.5.3).

Indeed, previous studies of the TF interaction with both isolated protein and nascent chain substrates found that the enrichment in aromatic residues was key to driving the interaction with TF [137]. The study of the TF interaction with isolated α-synuclein and ddFLN5 thus has consolidated at a residue specific level, that the major substrate interactions are in fact driven by such aromatic residues.

6.2.3 The structural investigation of TF on the ribosome by NMR spectroscopy

The initial characterisation of the ^{1}H-^{13}C HMQC spectrum of TF in the presence of the ribosome presented in Section 4.5 (Fig 4.10) suggested that TF might be observable while bound on the ribosome. This would constitute a significant feat since the ribosome is a 2.5 MDa particle, likely to induce rapid relaxation of the magnetisation of interacting nuclei. Indeed, TF would need to maintain significant flexibility to remain detectable by

NMR even in methyl TROSY based experiments with optimal selective methyl labelling. The dependence of NMR relaxation rates on the rotational correlation time as an indirect reporter of molecular size is discussed in Section 2.1.2.

The comparison of experimental resonance intensities in the presence of the ribosome (Section 4.5) to predicted intensities, in the case where free monomer and dimeric TF contribute to the observed signal only based on the deconvoluted spectra presented in Section 3.5.2.3, indicate the PPIase domain, and to a lesser extend the SBD domain of ribosome-bound TF contribute to the observed spectrum. This gradient in resonance intensity (Fig 4.10) from regions most distal to those nearer to the ribosome surface re-enforce the idea that the ribosome-bound species is indeed detected.

However, there remain a number of concerns that prevent the conclusive determination of the observability of the TF-ribosome bound state. The first corresponds to the differences in experimental conditions between the study of the TF interaction with the ribosome (Section 4.5), undertaken at 700 MHZ with H_6-TF^{ile} and the TF titration study (Section 3.5.2) on which the prediction about resonance intensities are based, 900 MHz with TF^{ILV}. Secondly, at 4°C significant broadening of TF^{ILV} ^{13}C signal at 700 MHz was observed (Section 4.6.3) while lower temperatures should result in slower dynamics and thus a more favourable (slow) exchange regime.

While literature values of 10 s for the half-life of the TF ribosome complex (summarised Table 3.1, [161]) would indicate that TF NMR resonances in the presence of the ribosome are in slow exchange as suggested by the results obtained at 25°C (Section 4.5), values for the half-life of the TF-ribosome complex of 100 ms, found in a more recent study [162], would appear to be more consistent with the observed intermediate exchange at 4°C (Section 4.6.3) in particular if chemical shift differences are relatively large. The presence of the H_6 tag on TF in the study undertake at 25°C (Section 4.5) is a possible source of the observed differences if the additional histidine residues form favourable interactions with the charged ribosome surface.

Further investigation of TF on the ribosome by NMR will therefore require careful consideration of experimental conditions: reducing the TF concentrations relative to the ribosome increase the population of the bound state and reduce back-ground contributions of the TF monomer and dimer observed in this study (Section 4.5) while careful selection of field strength and temperature may alter the exchange kinetics sufficiently to create a favourable slow exchange regime.

6.2.4 Residue specific analysis of the TF interaction with the α-synuclein and ddFLN5+21 RNCs

The disordered nascent chains, α-synuclein and ddFLN5+21 RNC, were shown to produce high quality NMR spectra (Section 4.6.1 and 5.4.2 respectively) despite challenges such as a maximum working ribosome concentration of 10 μM and limited sample lifetimes (∼ 20 hours at 10°C for the ddFLN5+21 RNC, Section 5.4.3). The absence of significant changes to resonance positions (Fig 4.11 and 5.5) indicated that the stalling of the nascent polypeptide on the ribosome did not induce major structural changes, while the resonance intensities relative to the isolated protein (10% and 40% for α-synuclein and ddFLN5+21 RNCs, Weise *et al.*, in preparation, and Section 5.5 respectively) indicate the tumbling of the α-synuclein and ddFLN5 nascent chains mirror that of flexible regions such as L7/L12 on the ribosome [183, 184].

The most C-terminal residue detected in α-synuclein RNC is D135 out of 161 residues including the SecM stalling sequence (Weise *et al.*, in preparation). This observation indicates that the nascent chains gains sufficient flexibility to be detected by NMR at approximately 25 residues from the PTC. The appearance of C-terminal resonances is delayed in the ddFLN5+21 RNC where residue 712, ∼60 removed from the PTC, is the most C-terminal residue observed with certainty (Fig 5.5, Section 5.4.2). With the ribosomal exit tunnel estimated to accommodate 30 residues [59, 260], the ddFLN5+21 RNC C-terminus appears to be in exchange either between conformational states or between free and ribosome-bound states.

The addition of excess concentrations of TF to the α-synuclein and ddFLN5+21 RNC results in significant broadening of both spectra to 50% and 65% respectively (Fig 4.11, Section 4.6.1 and Fig 5.7, Section 5.4.4) while no chemical shift perturbations are detected. The residue specific analysis of the resonance intensities shows that the broadening is uniform across the 100 N-terminal residues of both the α-synuclein and ddFLN5+21 RNC. Residues 120 to 135 of the α-synuclein RNC appear unperturbed by the presence of TF (Fig 4.11, Section 4.6.1). The equivalent residues in the ddFLN5+21 RNC (730-745) are not observed, even in the absence of TF (Section 5.4.2). Approximately 15-20 residues thus appear required to bridge the space between the exit tunnel where the α-synuclein RNC gains significant motional freedom as described above and the first possible contact with TF at the ribosome surface. A total NC length of 40-45 residues is thus required for first possible contact with TF on the ribosome. This is in agreement with previous cross-linking studies that found 47 residues were required before interaction with TF was observed [132].

The uniform broadening across the first 100 residues upon TF addition of both α-synuclein and ddFLN5 RNC resonances (Section 4.6.1 and 5.4.4) indicates no specific,

localised, interactions with TF occur. The similarity in the behaviour of both RNCs suggests a common phenomenon is detected. The uniform broadening of NMR resonances detected could be caused either by non specific interactions with TF across the nascent chain sequence and/or a steric hinderance by TF resulting in reduced motional freedom of the nascent chain (Fig 6.2). Both effects would be conducive to prevent aggregation and other non-native interactions in the emerging nascent chain suggesting that even in the absence of specific interaction sites, TF might play an important role in the cotranslational behaviour of nascent chains.

6.2.5 The mechanism of TF interaction with isolated protein substrates differs from that on the ribosome

The study of the TF interaction with α-synuclein in isolation (Section 4.4) and in the presence of the ribosome (Section 4.6) display significant differences. A weak N-terminal interaction with WT α-synuclein was identified in the absence of the ribosome (Fig 4.2). The high local concentrations of TF at the ribosome surface might suggest that this interaction would be amplified in the α-synuclein RNC. However, the effect of TF on the resonance intensities of α-synuclein RNC were uniform through-out the the majority of the sequence (Fig 4.11, Weise *et al.*, in preparation).

The study of the TF interaction with the ddFLN5+21 RNC (Section 5.4.4) mirrors the observations with the α-synuclein RNC. While no interaction was identified between TF and ddFLN5Δ16 (Fig 5.3, Section 5.3.2), a weak interaction was detected between TF and ddFLN5Δ12 (Section 5.5.3) indicating the region from residue 672 to 677 shows some propensity for interaction with TF. Again, the high local concentrations at the ribosome surface may amplify this interaction yet the observed resonance broadening in the ddFLN5+21 RNC in the presence of TF (Fig 5.7) was uniform. The detected transient interactions of TF with the isolated α-synuclein (Section 4.4) and ddFLN5Δ16 (Section 5.3.2) do not appear to play a role on the ribosome.

These early results of the TF interaction with nascent polypeptide chains studied at a residues specific level suggest that the TF interaction with substrates on the ribosome is distinct from that observed in isolation. A detailed characterisation of the bimolecular interaction is therefore not sufficient for the complete understanding of the TF chaperoning mechanism on the ribosome. Factors such as the high local concentrations of TF and the nascent chain, the charged ribosome surface and the tethering of the both TF and the nascent chain to the ribosome thus appear to modulate the interaction between TF and the nascent chain.

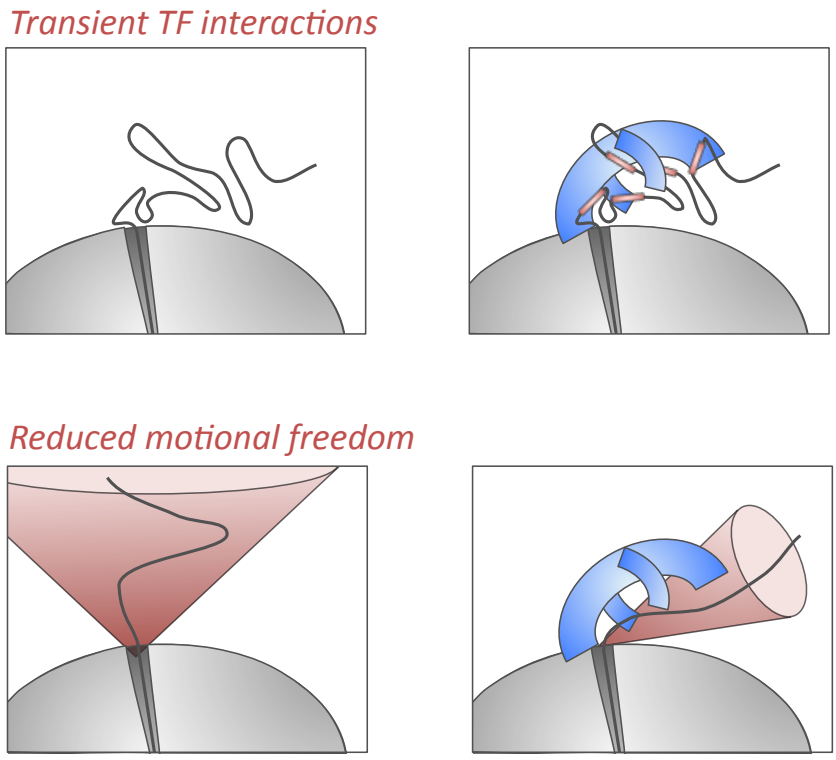

Figure 6.2: Proposed mechanisms of TF interaction with disordered nascent chains. The observed resonance broadening may reflect transient non specific interaction (pink)of TF (blue) with the nascent chain (grey) and/or report on the loss of rotational freedom (red cone) in the presence of TF.

6.2.6 Co^{2+} PREs as molecular ruler for compaction in the nascent chain

Chemical shift perturbations provide an excellent measure of stables secondary and tertiary structure while intensities are excellent probes of exchange phenomena. Here, we developed an approach to probe residual structure in disordered proteins and nascent chains. While typical PRE studies selectively introduce nitroxide spin labels on cysteine residues, the specific modification of nascent chains on the ribosome poses a great challenge.

The PRE arising from Co^{2+} binding to the H-terminal H_6-tag of ddFLN5 was explored as a means of detecting compaction and changes in the unfolded ensemble of the disordered ddFLN5+21 RNC in the presence of TF. Co^{2+} ions were found to interact with the disordered ddFLN5Δ16 with $\sim$10 μM affinity and induced broadening of resonances mapping to the ddFLN5Δ16 N-terminus up to 40 residues removed from the H_6-tag (Section 5.6.3). This broadening was characteristic to the disordered state since the folded ddFLN5 resonances were affected more locally, correlating with the native three dimensional structure. The presence of ribosomes did not disrupt the PRE pattern observed for ddFLN5Δ16 but elevated concentrations of Co^{2+} were required to obtain the same extent

of resonance broadening (Section 5.6.5). The 70S ribosome thus appears to bind Co^{2+} ions in possible competition with the native Mg^{2+}. A study of the ribosome associated, Co^{2+} dependent PDF found that the ribosome remains stable and active in the presence of Co^{2+} concentrations up to 100 μM [162] indicating that the study of Co^{2+} PRE of the ddFLN5+21 RNC is compatible with the ribosome.

A nitroxide spin label is predicted to induce PREs in a polypeptide in the random coil conformations up to 15 amino acids from its site of attachment [261]. The free electron of Co^{2+} ions has an electron relaxation rate three orders of magnitude greater that the nitroxide radical indicating that the PRE of Co^{2+} is predicted to have a significantly shorter reach. The observation that up to 40 residues from the ddFLN5Δ16 N-terminus are affected (Section 5.6.4) therefore suggests the ddFLN5 unfolded state adopts a compact ensemble in agreement with diffusion constants below predicted values (Marilia Karyadi, doctoral thesis).

6.3 Implications for the TF equilibrium *in vivo*

Taken together, the results presented in this thesis provide a detailed look in the complex mechanisms of TF/substrate interactions, and their fine-tuned interplay within the TF equilibrium. In particular, the dynamic insights into the TF dimer (Section 3.5.3.2) indicate this state may play a more prominent role in the cell than previously thought. The step-wise dissociation of the TF dimer provides the basis for a mechanism of substrate interaction regulation that does not rely entirely on the affinity of the substrate for the TF monomer state. Indeed, as the dimerisation interface largely overlaps with the substrate binding interface, the TF-substrate interaction needs to be of sufficient strength to outcompete the dimer interactions. This indicates that only substrates that can occupy multiple binding sites on TF and, as such, prevent the formation of both symmetrical dimer interfaces, can effectively remain associated with the TF monomer. The formation of a first dimer contact would drive the displacement of the substrate protein in a concerted and rapid release mechanism as the high local concentration of the TF dimer interface residues outcompetes the substrate interactions Fig 6.3).

In our study, TF displays a clear preference for aromatic residues. As such, sequences with TF affinity are likely to occur on a regular basis across polypeptide chains. It would be the initial folding and burial of one or more of these interaction sequences that would trigger the concerted release of the polypeptide permitting the rapid folding of the protein. This mechanism is equally applicable to isolated unfolded polypeptides and ribosome associated nascent chains in the absence of the stabilising interaction of the TF RBD at the ribosome surface which prevents TF dimerisation by burial of key contacts in the RBD in the ribosome interaction surface (Fig 6.3).

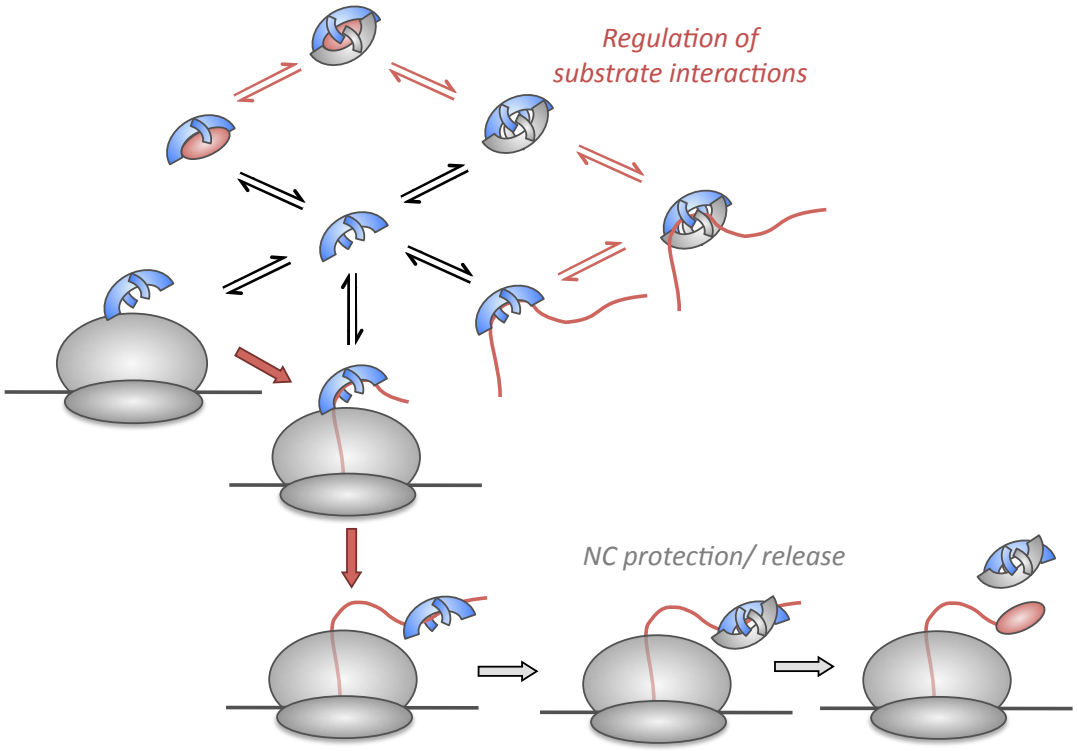

Figure 6.3: A more complex TF equilibrium? Reprisal of Fig 1.10 expanded to included additional TF-substrate bound states (pink equilibria) identified in this work and others, highlighting the more central role of the TF dimer. In grey arrows, the proposed mechanism for dimerisation-mediated nascent chain release.

The mathematical modelling of the TF distribution in the cell (Section 3.4.6) demonstrates that in the absence of competing cellular factors, nearly all active ribosomes can be occupied by TF due to the relative higher affinity for such types of substrates, stressing the importance of the TF role at the ribosome exit tunnel. Interestingly, the accumulation of isolated high affinity TF substrates in the cytosol results in the interconversion of the TF dimeric pool to a substrate bound state without affecting the extend of ribosome-association. The excess dimer TF thus serves not only as a storage form to prevent unwanted interactions, it provides an active buffer against the accumulation of unfolded, aggregation-prone species in the cytosol, while ensuring that its housekeeping activity on the ribosome is unaffected.

The potential ubiquitous presence of TF at the exit tunnel, even in the absence of high affinity sequences in the nascent chain was shown in Sections 4.6.1 and 5.4.4 to significantly affect the conformational space of these emerging polypeptides. TF chaperoning function is therefore not entirely reliant on interaction, but its mere presence at the exit tunnel may shield and confine the nascent chain until folding occurs. Overall, the results presented in this thesis paint a picture of a highly abundant, multi-functional chaperone in the *E. coli*

cytosol whose ATP independent activity appears to rely on a complex interplay between desirable substrate interactions.

6.4 Concluding remarks

The study of the structural and dynamic aspects of the various TF equilibria described in this thesis has opened an number of avenues for further investigation. While this thesis characterises the interaction of TF with nascent chains that do not display high affinity TF interaction sites, the foundations, in the form of a preliminary characterisation of the α-syn (Luc 87-100) and ddFLN5Δ16 tryptophan mutant F672W, have been laid to expand the current study to include specific interactions.

Moreover, the prospect of investigating TF bound to the ribosome in structural detail is exciting as the high resolution structure of the full-length TF has thus far remained elusive to other techniques. Indeed, its conformations on the ribosome may dictate the way in which it interacts with its candidate substrates, and its effect on the conformational space sampled by emerging nascent chains. Finally, we have developed a number of mathematical approaches and NMR protocols that are readily applicable to highly detailed investigation of dynamic processes in other large biological systems.

Appendix A

A.1 DNA and protein sequences

A.1.1 Trigger factor

Trigger factor DNA sequence

ATG CAA GTT TCA GTT GAA ACC ACT CAA GGC CTT GGC CGC CGT GTA
ACG ATT ACT ATC GCT GCT GAC AGC ATC GAG ACC GCT GTT AAA AGC
GAG CTG GTC AAC GTT GCG AAA AAA GTA CGT ATT GAC GGC TTC CGC
AAA GGC AAA GTG CCA ATG AAT ATC GTT GCT CAG CGT TAT GGC GCG
TCT GTA CGC CAG GAC GTT CTG GGT GAC CTG ATG AGC CGT AAC TTC
ATT GAC GCC ATC ATT AAA GAA AAA ATC AAT CCG GCT GGC GCA CCG
ACT TAT GTT CCG GGC GAA TAC AAG CTG GGT GAA GAC TTC ACT TAC
TCT GTA GAG TTT GAA GTT TAT CCG GAA GTT GAA CTC GAG GGT CTG
GAA GCG ATC GAA GTT GAA AAA CCG ATC GTT GAA GTG ACC GAC GCT
GAC GTT GAC GGC ATG CTG GAT ACT CTG CGT AAA CAG CAG GCG ACC
TGG AAA GAA AAA GAC GGC GCT GTT GAA GCA GAA GAC CGC GTA ACC
ATC GAC TTC ACC GGT TCT GTA GAC GGC GAA GAG TTC GAA GGC GGT
AAA GCG TCT GAT TTC GTA CTG GCG ATG GGC CAG GGT CGT ATG ATC
CCG GGC TTT GAA GAC GGT ATC AAA GGC CAC AAA GCT GGC GAA GAG
TTC ACC ATC GAC GTG ACC TTC CCG GAA GAA TAC CAC GCA GAA AAC
CTG AAA GGT AAA GCA GCG AAA TTC GCT ATC AAC CTG AAG AAA GTT
GAA GAG CGT GAA CTG CCG GAA CTG ACT GCA GAA TTCA TCA AAC GTT
TCG GCG TTG AAG ATG GTT CCG TAG AAG GTC TGC GCG CTG AAG TGC
GTA AAA ACA TGG AGC GCG AGC TGA GAG CGC CAT CCG TAA CCG CGT
TAA GTT CTC AGG CGA TCG AAG GTC TGG TAA AAG CTA ACG ACA TCG
ACG TAC CGG CTG CGC TGA TCG ACA GCG AAA TCG ACG TTC TGC GTC
GCC AGG CTG CAC AGC GTT TCG GTG GCA ACG AAA AAC AAG CTC TGG
AAC TGC CGC GCG AAC TGT TCG AAG AAC AGG CTA AAC GCC GCG TAG
TTG TTG GCC TGC TGC TGG GCG AAG TTA TCC GCA CCA ACG AGC TGA

AAG CTG ACG AAG AGC GCG TGA AAG GCC TGA TCG AAG AGA TGG CTT
CTG CGT ACG AAG ATC CGA AAG AAG TTA TCG AGT TCT ACA GCA AAA
ACA AAG AAC TGA TGG ACA ACA TGC GCA ATG TTG CTC TGG AAG AAC
AGG CTG TTG AAG CTG TAC TGG CGA AAG CGA AAG TGA CTG AAA AAG
AAA CCA CTT TCA ACG AGC TGA TGA ACC AGC AGG CGT AAT TTA CGC
AGC ATA ACG CGC TAA ATT CGC ACA AAG GCC CGT CAC CGC CAG GTG
GTG GGC TTT TTT TTG TCA TGA ATT TTG CAT GGA ACC GTG CGA AAA
GCC TCT TTC GGT GTT AGC GTA ACA ACA AAA GAT TGT TAT GCT TGA
AAT ATG GTG ATG CCG TAC CCA TAA CAC AGG GAC TAG CTG ATA ATC
CGT CCA TAA GGT TAC AAT CGG TAC AGC AGG TTT TTT CAA TTT TAT
CCA GGA GAC GGA AAT GTC ATA CAG CGG CGA ACG AGA TAA CTT TGC
ACC CCA TAT GGC GCT GGT GCC GAT GGT CAT TGA ACA GAC CTC CAC
GCG GTG AGC

Trigger factor protein sequence

cysteine mutations in red - R14C, T150C, N325C and M376C

Q V S V E T T Q G L G R R V T I T I A A D S I E T A V K S E L V N V A K K V R I
D G F R K G K V P M N I V A Q R Y G A S V R Q D V L G D L M S R N F I D A I I
K E K I N P A G A P T Y V P G E Y K L G E D F T Y S V E F E V Y P E V E L E G
L E A I E V E K P I V E V T D A D V D G M L D T L R K Q Q A T W K E K D G A
V E A E D R V T I D F T G S V D G E E F E G G K A S D F V L A M G Q G R M I P
G F E D G I K G H K A G E E F T I D V T F P E E Y H A E N L K G K A A K F A I
N L K K V E E R E L P E L T A E F I K R F G V E D G S V E G L R A E V R N M E
R E L R A P S V T A L S S Q A I E G L V K A N D I D V P A A L I D S E I D V L R
R Q A A Q R F G G N E K Q A L E L P R E L F E E Q A K R R V V V G L L L G E
V I R T N E L K A D E E R V K G L I E E M A S A Y E D P K E V I E F Y S K N K E
L M D N M R N V A L E E Q A V E A V L A K A K V T E K E T T F N E L M N Q Q A

Primers TF C14R

Forward : GGCCTTGGCCGCCGTGTAACGATTACT

Reverse : AGTAATCGTTACACGGCGGCCAAGGCC

A.1.2 α-synuclein

α-synuclein DNA sequence

His-tag in blue

ATG CAT CAC CAT CAC CAT CAC GCT AGC ATG GAT GTA TTC ATG AAA GGA
CTT TCA AAG GCC AAG GAG GGA GTT GTG GCT GCT GCT GAG AAA ACC
AAA CAG GGT GTG GCA GAA GCA GCA GGA AAG ACA AAA GAG GGT GTT
CTC TAT GTA GGC TCC AAA ACC AAG GAG GGA GTG GTG CAT GGT GTG
GCA ACA GTG GCT GAG AAG ACC AAA GAG CAA GTG ACA AAT GTT GGA
GGA GCA GTG GTG ACG GGT GTG ACA GCA GTA GCC CAG AAG ACA GTG
GAG GGA GCA GGG AGC ATT GCA GCA GCC ACT GGC TTT GTC AAA AAG
GAC CAG TTG GGC AAG AAT GAA GAA GGA GCC CCA CAG GAA GGA ATT
CTG GAA GAT ATG CCT GTG GAT CCT GAC AAT GAG GCT TAT GAA ATG
CCT TCT GAG GAA GGG TAT CAA GAC TAC GAA CCT GAA GCC

α-synuclein protein sequence

H H H H H H A S M D V F M K G L S K A K E G V V A A A E K T K Q G V A E A
A G K T K E G V L Y V G S K T K E G V V H G V A T V A E K T K E Q V T N V
G G A V V T G V T A V A Q K T V E G A G S I A A A T G F V K K D Q L G K N
E E G A P Q E G I L E D M P V D P D N E A Y E M P S E E G Y Q D Y E P E A

α-syn (Luc 87-100) DNA sequence

Firefly luciferase residues 87-100 substitution in red

ATG CAT CAC CAT CAC CAT CAC GCT AGC ATG GAT GTA TTC ATG AAA GGA
CTT TCA AAG GCC AAG GAG GGA GTT GTG GCT GCT GCT GAG AAA ACC
AAA CAG GGT GTG GCA GAA GCA GCA GGA AAG ACA AAA GAG GGT GTT
CTC TAT GTA GGC TCC AAA ACC AAG GAG GGA GTG GTG CAT GGT GTG
GCA ACA GTG GCT GAG AAG ACC AAA GAG CAA GTG ACA AAT GTT GGA
GGA GCA GTG GTG ACG GGT GTG ACA GCA GTA GCC CAG AAG ACA GTG
GAG GGA GCA GGG CAA TTC TTT ATG CCG GTG TTG GGC GCG TTA TTT
ATC GGA GTT GGC AAG AAT GAA GAA GGA GCC CCA CAG GAA GGA ATT
CTG GAA GAT ATG CCT GTG GAT CCT GAC AAT GAG GCT TAT GAA ATG
CCT TCT GAG GAA GGG TAT CAA GAC TAC GAA CCT GAA GCC

α-syn (Luc 87-100) protein sequence

H H H H H H A S M D V F M K G L S K A K E G V V A A A E K T K Q G V A E A
A G K T K E G V L Y V G S K T K E G V V H G V A T V A E K T K E Q V T N V
G G A V V T G V T A V A Q K T V E G A G Q F F M P V L G A L F I G V G K N
E E G A P Q E G I L E D M P V D P D N E A Y E M P S E E G Y Q D Y E P E A

Primers α-syn (Luc 87-100)
Forward : GGTGTGACAGCAGTACAATTCTTTATGCCGGTGTTGGGCGCGTTATT-TATCGGAGTTACTGGCTTTGTCAAA
Reverse : TTTGACAAAGCCAGTAACTCCGATAAATAACGCGCCCAACACCGGCATAAA-GAATTGTACTGCTGTCACACC

A.1.3 ddFLN5

ddFLN5-6δ18 DNA sequence
His-tag in blue
ATG CAT CAC CAT CAC CAT CAC GCT AGC AAA CCA GCC CCA TCC GCT GAA CAC TCT TAT GCT GAA GGT GAA GGT TTA GTC AAA GTA TTT GAT AAT GCC CCA GCT GAA TTC ACT ATT TTC GCC GTT GAC ACT AAA GGT GTT GCT CGT ACC GAT GGT GGT GAT CCA TTT GAA GTC GCT ATC AAT GGT CCA GAT GGT TTA GTC GTT GAT GCC AAA GTT ACC GAT AAC AAT GAC GGT ACT TAT GGT GTT GTC TAT GAT GCC CCA GTT GAA GGT AAC TAC AAT GTT AAT GTC ACC CTC CGT GGT AAT CCA ATC AAA AAT ATG CCA ATC GAT GTC AAA TGC ATT GAA GGT GCC AAT GGT GAA GAT TCA TCA TTC GGT TCA TTC ACT TTT ACC GTC GCT GCT AAA AAT AAG AAA GGT GAA GTT AAA ACC TAT GGT GGT GAT AAA TTT GAA GTC TCT ATC ACT GGT CCA GCT GAA GAA ATC ACT CTC GAT GCT ATT GAT AAC CAA GAT GGT ACT TAT ACT GCC GCT TAC TCT TTA GTT GGT AAT GGT CGT TTC TCA ACT GGT GTC AAA TTA AAC GGT AAA CAC ATT GAA GGT TCT CCA TTC AAA CAA GTA CTT GGT AAC TAA

ddFLN5+21 RNC DNA sequence
SecM sequence in red
ATG CAT CAC CAT CAC CAT CAC GCT AGC AAA CCA GCC CCA TCC GCT GAA CAC TCT TAT GCT GAA GGT GAA GGT TTA GTC AAA GTA TTT GAT AAT GCC CCA GCT GAA TTC ACT ATT TTC GCC GTT GAC ACT AAA GGT GTT GCT CGT ACC GAT GGT GGT GAT CCA TTT GAA GTC GCT ATC AAT GGT CCA GAT GGT TTA GTC GTT GAT GCC AAA GTT ACC GAT AAC AAT GAC GGT ACT TAT GGT GTT GTC TAT GAT GCC CCA GTT GAA GGT AAC TAC AAT GTT AAT GTC ACC CTC CGT GGT AAT CCA ATC AAA AAT ATG CCA ATC GAT GTC AAA TGC ATT GAA GGT ACT AGT GAA TTC TTC AGC ACG CCC GTC TGG ATA AGC CAG GCG CAA GGC ATC CGT GCT GGC CCT TAA

ddFLN5Δ16 protein sequence

His-tag in blue, tryptophan substitutions in cyan (Y655W, F665W, A668W, F672W and E724W)

H H H H H H A S K P A P S A E H S Y A E G E G L V K V F D N A P A E F T I F
A V D T K G V A R T D G G D P F E V A I N G P D G L V V D A K V T D N N D
G T Y G V V Y D A P V E G N Y N V N V T L R

ddFLN5Δ6 protein sequence

H H H H H H A S K P A P S A E H S Y A E G E G L V K V F D N A P A E F T I F
A V D T K G V A R T D G G D P F E V A I N G P D G L V V D A K V T D N N D
G T Y G V V Y D A P V E G N Y N V N V T L R G N P I K N M P I D

ddFLN5 protein sequence

H H H H H H A S K P A P S A E H S Y A E G E G L V K V F D N A P A E F T I F
A V D T K G V A R T D G G D P F E V A I N G P D G L V V D A K V T D N N D
G T Y G V V Y D A P V E G N Y N V N V T L R G N P I K N M P I D V K C I E G

ddFLN5-6δ18 protein sequence

ddFLN6 sequence in green

H H H H H H A S K P A P S A E H S Y A E G E G L V K V F D N A P A E F T I F
A V D T K G V A R T D G G D P F E V A I N G P D G L V V D A K V T D N N D
G T Y G V V Y D A P V E G N Y N V N V T L R G N P I K N M P I D V K C I E G
A N G E D S S F G S F T F T V A A K N K K G E V K T Y G G D K F E V S I T G
P A E E I T L D A I D N Q D G T Y T A A Y S L V G N G R F S T G V K L N G K
H I E G S P F K Q V L G N

ddFLN5+21 RNC protein sequence

SecM sequence in red

H H H H H H A S K P A P S A E H S Y A E G E G L V K V F D N A P A E F T I F
A V D T K G V A R T D G G D P F E V A I N G P D G L V V D A K V T D N N D
G T Y G V V Y D A P V E G N Y N V N V T L R G N P I K N M P I D V K C I E G
T S E F F S T P V W I S Q A Q G I R A G P

Primers Y655W

Forward :TCC GCT GAA CAC TCT TGG GCT GAA GGT GAA GGT

Reverse :ACCTTCACCTTCAGCCCAAGAGTGTTCAGCGGA

Primers F665W
Forward :GGT TTA GTC AAA GTA TGG GAT AAT GCC CCA GCT
Reverse :AGCTGGGGCATTATCCCATACTTTGACTAAACC

Primers A668W
Forward :AAA GTA TTT GAT AAT TGG CCA GCT GAA TTC ACT
Reverse :AGTGAATTCAGCTGGCCAATTATCAAATACTTT

Primers F672W
Forward :AAT GCC CCA GCT GAA TGG ACT ATT TTC GCC GTT
Reverse :AACGGCGAAAATAGTCCATTCAGCTGGGGCATT

Primers E724W
Forward :TAT GAT GCC CCA GTT TGG GGT AAC TAC AAT GTT
Reverse :AACATTGTAGTTACCCCAAACTGGGGCATCATA

A.2 NMR processing scripts

```
nmrPipe -in test.fid \
| nmrPipe -fn EM  -lb 12.0 -c 1.0           \
| nmrPipe  -fn ZF -auto                              \
| nmrPipe  -fn FT -auto                               \
| nmrPipe  -fn PS -p0 -35 -p1 0.00 -di -verb         \
| nmrPipe -fn EXT -left -sw               \
| nmrPipe  -fn TP                                     \
| nmrPipe  -fn SP -off 0.5 -end 1 -pow 1 -c 1.0    \
| nmrPipe  -fn ZF -auto                              \
| nmrPipe  -fn FT -auto                               \
| nmrPipe  -fn PS -p0 -90.00 -p1 180.00 -di -verb      \
   -ov -out test.ft2
```

Figure A.1: Representative NMRPipe processing script with an exponential window function for RNC samples.

```
xyz2pipe -in fid/test%03d.fid -x -verb \
| nmrPipe -fn GM  -g1 8.0 -g2 24 -g3 0.0 -c 0.5              \
| nmrPipe  -fn ZF -auto                              \
| nmrPipe  -fn FT                                     \
| nmrPipe  -fn PS -p0 42 -p1 0.0 -di              \
| nmrPipe  -fn EXT -xn 1.5ppm -x1 -1ppm -sw                        \
| nmrPipe  -fn TP \
| nmrPipe -fn GM  -g1 8.0 -g2 18.0 -g3 0.0 -c 1.0              \
| nmrPipe  -fn ZF -auto                              \
| nmrPipe  -fn FT -alt -neg                                    \
| nmrPipe  -fn PS -p0 -90.0 -p1 180.0 -di               \
| pipe2xyz -out ft/test%03d.ft2 -ov -verb
```

Figure A.2: Representative NMRPipe processing script with a Gaussian window function for TF^{ILV} relaxation measurements .

```
nmrPipe -in test.fid \
| nmrPipe  -fn SP -off 0.5 -end 1.00 -pow 1 -c 0.5    \
| nmrPipe  -fn ZF -auto                               \
| nmrPipe  -fn FT -auto                               \
| nmrPipe  -fn PS -p0 142 -p1 0.00 -di -verb         \
   -ov -out test.ft2
```

Figure A.3: Representative NMRPipe processing script with a sine-bell window function for 1D NMR experiments.

A.3 Matlab scripts

```
clear all

%% Load experimental data
%loadSpectra; % creates data{} and dN, dH vectors

load '/Users/Anne/Desktop/chris_tf_conc_130214/5buM.txt'
load '/Users/Anne/Desktop/chris_tf_conc_130214/10buM.txt'
load '/Users/Anne/Desktop/chris_tf_conc_130214/15buM.txt'
load '/Users/Anne/Desktop/chris_tf_conc_130214/20buM.txt'
load '/Users/Anne/Desktop/chris_tf_conc_130214/30buM.txt'
load '/Users/Anne/Desktop/chris_tf_conc_130214/50buM.txt'
load '/Users/Anne/Desktop/chris_tf_conc_130214/75buM.txt'

dC=X5buM(1:256,1);
dH=X5buM(1:256:end,2);

noise=1e5;

data{1}=reshape(X5buM(:,3),[length(dC) length(dH)])/noise;
data{2}=reshape(X10buM(:,3),[length(dC) length(dH)])/noise;
data{3}=reshape(X15buM(:,3),[length(dC) length(dH)])/noise;
data{4}=reshape(X20buM(:,3),[length(dC) length(dH)])/noise;
data{5}=reshape(X30buM(:,3),[length(dC) length(dH)])/noise;
data{6}=reshape(X50buM(:,3),[length(dC) length(dH)])/noise;
data{7}=reshape(X75buM(:,3),[length(dC) length(dH)])/noise;

%% normalise spectra

ns = [160 80 32 20 16 12 8];
for i=1:7
    data{i} = data{i} / ns(i);
end

%% determine noise in spectra
for i=1:7
    d=data{i};
    d=d(210:250,30:50);
    noise(i)=std(d(:));
end

%% convert data to signal to noise

for i=1:7
    data_snr{i}= data{i} / noise(i);
end


%% plot the observed data

for i=1:length(data)
    %mask=data{i}/conc(i) >0.1;
    mask=data_snr{i} >10;
    subplot(2,4,i)
    %imagesc(dH,dC,data_snr{i})
    imagesc(dH,dC,mask)
    set(gca,'xdir','reverse');
    set(gca,'ydir','reverse');
    %xlim(xl);ylim(yl);
    ax(i)=gca;
end
linkaxes(ax);

%% roll up spectra into matrix
vec = @(x)x(:);
for i=1:7 %8
    Yobs(:,i) = vec(data{i});
    Ysnr(:,i) = vec(data_snr{i});
end

idx = max(Ysnr,[],2)>10; % threshold data
Ysnr = Ysnr(idx,:);
```

continued ...

... continued

```
%% calculate expected monomer/dimer concentrations
%Kd = 11;
for Kd=8:0.2:14

conc = [4 9 13 18 26 43 65]; % 100];

M = 0.25*(-Kd+sqrt(Kd^2+8*conc*Kd));
D = 0.5*(conc - M);

Msnr = M ./ noise;
Dsnr = D ./ noise;
D2snr = 2*Dsnr; % monomer in form of dimer

%plot(conc,Msnr,'ob-',conc,Dsnr,'rs-')

%% do the regression analysis


X = [Msnr; D2snr];
X2=X';

Yobs2=Ysnr';
clear Y;
chi2=0.;
for i=1:length(Yobs2)
    [Y(:,i),resnorm]=lsqnonneg(X2,Yobs2(:,i));
    chi2=chi2+resnorm;
end
Y=Y';

Ym = Y(:,1);
Yd = Y(:,2);


fprintf(' Kd= %f\t chi2= %f\n',Kd, chi2);

end


%% plot the resolved spectra

specM = reshape(Ym, [length(dC) length(dH)]);
specD = reshape(Yd, [length(dC) length(dH)]);


subplot(1,3,1)
imagesc(dH,dC,data{7})
set(gca,'xdir','reverse');
set(gca,'ydir','reverse');
title('100 uM spectrum')
ax(1)=gca;

subplot(1,3,2)
%imagesc(dH,dC,specM)
hold on
contour(dH,dC,specM,10.^linspace(-1.7,0,10),'r')
hold off
set(gca,'xdir','reverse');
set(gca,'ydir','reverse');
title('Monomer')
ax(2)=gca;
caxis([0 .4])

subplot(1,3,3)
%imagesc(dH,dC,specD)
hold on
contour(dH,dC,specD,10.^linspace(-1.7,0,10),'r')
hold off
set(gca,'xdir','reverse');
set(gca,'ydir','reverse');
title('Dimer')
ax(3)=gca;
caxis([0 .4])

linkaxes(ax);
```

Figure A.4: Matlab script for the linear, non-negative, regression fit of the monomer/dimer equilibria to the total NMR spectrum.

```
function residuals=calcChi2_syn(x1,x2, yobs2, yobs3, param)

%calculate residuals for:
%15N data: y1=1-(solve_b4(1.35,K2,K3,6.5,x,0.8)/6.5)+c*(solve_b4(1.35,2,K3,6.5,x,0.8)/6.5)
%13C 70S data y2=1e7*a*solve_m4(1.35,K2,K3,5,x,0)+1e7*b*(solve_m4(1.35,K2,K3,5,x,0)/1.35)
%13C RNC data y3=1e7*a*solve_m4(1.35,K2,K3,6.5,x,0.8)+1e7*b*(solve_m4(1.35,K2,K3,6.5,x,0.8)/1.35)
% a: Intensity TF monomer = param(3)
% b: Intensity TF dimer = param(4)
% c: Intensity bound RNC= param(5)
% K2: ribosome binding Kd = param(1)
% K3: RNC binding Kd = param(2)
% x1: set of TF concentrations in RNC experiment
%x2: set of TF concentrations in 70S titration

% solve_m4 is the numerical solution to the four state TF equlibrium in the presence of the RNC for the free TF
monomer concentration as a function of Kd for dimerisation, Kd for ribosome interaction, Kd for RNC
interaction , total ribosome concentration, total TF concentration and ribosome occupancy.

ycalc2=1e7*param(3)*solve_m4(8,param(1),param(2),5,x2,0)+1e7*param(4)*(solve_m4(8,param(1),param(2),
5,x2,0).^2/8);
ycalc3=1e7*param(3)*solve_m4(8,param(1),param(2),6.5,x1,0.8)+1e7*param(4)*(solve_m4(8,param(1),param(2),
6.5,x1,0.8).^2/8);
residuals2=ycalc2-yobs2;
residuals3=ycalc3-yobs3;
residuals=[residuals2 residuals3];
```

Figure A.5: Matlab script for global fit of the TF^{ILV} titrations of 70S and α -synuclein RNC.

A.4 α-synuclein (Luc 87-100) NMR backbone assignment

Residue name	**Residue type**	**H**	**N**	**Co**	**Cα**	**Cβ**
1	Met	8.474	122.289	177.575	55.818	32.632
2	Asp	8.325	121.456	177.046	55.194	40.994
3	Val	8.001	119.612	177.176	63.546	32.285
4	Phe	8.27	122.754	177.421	58.758	39.216
5	Met	8.329	120.958	177.169	55.858	32.225
6	Lys	8.288	122.024	176.388	57.305	32.706
7	Gly	8.373	109.335	179.45	45.489	-
8	Leu	8.039	121.582	176.051	55.344	42.373
9	Ser	8.357	116.536	179.446	58.707	63.634
10	Lys	8.425	123.73	177.271	56.411	32.903
11	Ala	8.369	125.359	175.825	52.805	19.046
12	Lys	8.425	121.004	177.085	56.587	32.943
13	Glu	8.517	122.338	176.765	56.882	30.182
14	Gly	8.554	110.217	179.839	45.325	-
15	Val	8.079	120.534	177.293	62.781	32.718
16	Val	8.412	125.682	177.773	62.613	32.741
17	Ala	8.565	128.542	176.073	52.619	19.078
18	Ala	8.428	123.802	175.875	52.837	18.961
19	Ala	8.38	123.305	175.564	52.829	18.998
20	Glu	8.429	120.391	176.875	56.886	30.176
21	Lys	8.441	122.552	176.619	56.836	32.865
22	Thr	8.255	115.762	179.158	62.505	69.791
23	Lys	8.455	123.929	177.058	56.737	32.905
24	Gln	8.541	122.061	177.218	55.781	29.398
25	Gly	8.576	110.679	179.556	45.347	-
26	Val	8.129	120.177	177.382	62.812	32.677
27	Ala	8.511	127.477	175.611	47.858	18.928
28	Glu	8.513	120.763	177.12	56.938	30.152
29	Ala	8.4	125.087	176.062	52.803	19.015
30	Ala	8.333	123.184	175.363	52.884	19.043
31	Gly	8.403	107.9	179.714	45.343	-

32	Lys	8.194	120.86	176.836	56.267	33.261
33	Thr	8.345	115.942	179.218	62.03	70.004
34	Lys	8.595	124.155	177.367	56.604	32.921
35	Glu	8.566	122.431	176.938	56.968	30.262
36	Gly	8.551	110.186	179.936	45.277	-
37	Val	8.023	119.871	177.926	62.377	32.722
38	Leu	8.404	126.226	177.228	54.964	42.416
39	Tyr	8.401	122.893	178.325	58.017	38.819
40	Val	8.189	123.947	177.815	62.188	32.795
41	Gly	8.117	112.395	180.014	45.106	-
42	Ser	8.362	115.712	179.133	58.433	63.967
43	Lys	8.596	123.656	177	56.318	33.147
44	Thr	8.301	115.888	179.298	61.955	69.963
45	Lys	8.565	124.109	177.374	56.617	32.943
46	Glu	8.554	122.146	176.907	57.002	30.108
47	Gly	8.486	109.95	180.084	45.091	-
48	Val	8.023	120.296	177.893	62.312	32.812
49	Val	8.434	125.619	178.033	62.172	32.738
50	His	8.612	125.34	178.014	56.449	31.096
51	Gly	8.51	110.814	180.159	45.147	-
52	Val	8.146	119.85	177.929	62.057	32.9
53	Ala	8.611	128.58	176.032	52.441	19.225
54	Thr	8.357	115.344	179.362	61.969	69.995
55	Val	8.381	123.625	178.017	62.333	32.873
56	Ala	8.55	128.494	176.077	52.539	19.221
57	Glu	8.495	121.222	177.171	56.613	30.344
58	Lys	8.564	123.257	176.888	56.524	33.01
59	Thr	8.342	116.508	179.272	62.264	69.822
60	Lys	8.516	124.036	177.162	56.688	32.985
61	Glu	8.543	122.485	177.44	56.906	32.854
62	Gln	8.54	122.102	177.907	55.789	29.528
63	Val	8.412	122.355	177.524	62.469	32.778
64	Thr	8.426	118.541	179.886	61.866	69.902
65	Asn	8.636	122.081	178.629	-	-
66	Val	8.367	121.078	176.978	62.776	32.482
67	Gly	8.674	112.905	179.241	45.278	-
68	Gly	8.329	108.947	180.214	44.985	-
69	Ala	8.269	123.84	176.219	52.293	19.289

70	Val	8.351	120.93	177.528	62.475	32.694
71	Val	8.533	126.013	177.595	62.153	32.747
72	Thr	8.44	119.173	179.01	61.961	69.939
73	Gly	8.551	111.592	179.941	45.242	-
74	Val	8.202	119.792	177.321	62.362	32.855
75	Thr	8.428	119.6	179.862	62.078	69.79
76	Ala	8.502	127.79	176.328	52.428	19.24
77	Val	8.285	120.569	177.865	62.261	32.816
78	Ala	8.54	128.549	176.244	52.511	19.059
79	Gln	8.519	120.632	177.947	55.682	29.594
80	Lys	8.555	123.619	177.189	56.372	33.147
81	Thr	8.408	117.29	179.46	61.99	69.947
82	Val	8.426	123.401	177.733	62.341	32.811
83	Glu	8.687	125.597	176.832	56.903	30.15
84	Gly	8.606	111.01	179.934	45.293	-
85	Ala	8.292	123.696	175.489	52.686	19.276
86	Gly	8.557	108.309	179.833	45.246	-
87	Gln	8.209	119.685	178.489	55.807	29.533
88	Phe	8.322	121.215	178.989	57.585	39.742
89	Phe	8.2	122.607	179.286	57.477	39.896
90	Met	8.274	124.491	180.47	-	-
91	Pro	-	-	177.295	62.913	32.143
92	Val	8.417	121.491	177.504	62.372	32.648
93	Leu	8.536	127.047	176.082	55.319	42.217
94	Gly	8.564	110.484	180.13	45.146	-
95	Ala	8.222	123.675	176.249	52.578	19.356
96	Leu	8.294	121.026	177.008	55.068	42.357
97	Phe	8.349	121.703	178.579	57.714	39.621
98	Ile	8.132	124.036	177.802	60.952	38.737
99	Gly	8.094	112.65	180.149	45.11	-
100	Val	8.197	119.129	177.072	62.265	32.763
101	Gly	8.704	113.028	178.623	45.164	-
102	Lys	8.354	121.028	177.525	56.186	33.224
103	Asn	8.72	120.244	178.612	53.294	38.731
104	Glu	8.594	121.581	177.325	56.62	30.175
105	Glu	8.567	122.187	176.826	56.901	30.21
106	Gly	8.537	110.473	180.494	44.974	-

107	Ala	8.223	125.095	178.344	50.53	18.056
108	Pro	-	-	176.87	63.079	32.098
109	Gln	8.71	121.503	177.911	55.665	29.635
110	Glu	8.65	122.941	177.041	56.669	30.471
111	Gly	8.598	110.532	180.178	45.21	-
112	Ile	8.12	120.386	177.587	60.954	38.605
113	Leu	8.531	127.412	176.711	54.993	42.311
114	Glu	8.542	122.499	178.008	56.44	30.425
115	Asp	8.476	121.722	178.082	54.277	41.059
116	Met	8.384	122.291	176.898	53.187	32.247
117	Pro	-	-	177.187	62.925	32.185
118	Val	8.439	121.252	178.098	61.996	33.046
119	Asp	8.659	126.398	179.166	52.077	41.053
120	Pro	-	-	176.996	63.507	32.165
121	Asp	8.474	119.427	177.753	54.529	40.839
122	Asn	8.193	119.352	178.467	53.484	39.238
123	Glu	8.492	122.077	177.78	56.896	30.116
124	Ala	8.345	124.569	176.638	52.361	19.079
125	Tyr	8.144	120.269	178.59	57.898	38.967
126	Glu	8.198	124.265	178.571	55.568	30.76
127	Met	8.528	124.189	179.609	53.342	32.194
128	Pro	-	-	177.038	63.086	32.227
129	Ser	8.599	116.993	179.091	58.318	63.947
130	Glu	8.692	123.414	177.361	56.606	30.137
131	Glu	8.569	122.175	176.902	56.767	30.137
132	Gly	8.536	110.283	180.054	45.108	-
133	Tyr	8.165	120.493	178.114	58.261	38.747
134	Gln	8.321	123.038	179.087	55.385	29.765
135	Asp	8.329	121.802	178.342	54.22	41.126
136	Tyr	8.157	120.777	178.786	57.687	39.013
137	Glu	8.342	125.69	180.266	53.603	30.152
138	Pro	-	-	177.048	62.89	32.184
139	Glu	8.635	121.838	178.46	56.565	30.289
140	Ala	8.134	107.792	176.937	53.865	20.062

Bibliography

[1] Hapke, B. & Noll, H. Structural dynamics of bacterial ribosomes. IV. Classification of ribosomes by subunit interaction. *J Mol Biol* **105**, 97–109 (1976).

[2] Bartlett, A.I. & Radford, S.E. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat Struct Mol Biol* **16**, 582–588 (2009).

[3] Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).

[4] Koshland, D.E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *P Natl Acad Sci USA* **44**, 98–104 (1958).

[5] Rosenbaum, D.M., Rasmussen, S.G.F. & Kobilka, B.K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).

[6] Dill, K.A. Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155 (1990).

[7] Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).

[8] Makhatadze, G.I. & Privalov, P.L. Hydration effects in protein unfolding. *Biophys. Chem.* **51**, 291–304– discussion 304–9 (1994).

[9] Makhatadze, G.I. & Privalov, P.L. Energetics of protein structure. *Adv. Protein Chem.* **47**, 307–425 (1995).

[10] Matouschek, A., Kellis, J.T., Serrano, L., Bycroft, M. & Fersht, A.R. Transient folding intermediates characterized by protein engineering. *Nature* **346**, 440–445 (1990).

[11] Otzen, D.E., Itzhaki, L.S., elMasry, N.F., Jackson, S.E. & Fersht, A.R. Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *P Natl Acad Sci USA* **91**, 10422–10425 (1994).

[12] Radford, S.E. & Dobson, C.M. Insights into protein folding using physical techniques: studies of lysozyme and alpha-lactalbumin. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **348**, 17–25 (1995).

[13] Karplus, M. & Weaver, D.L. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.* **3**, 650–668 (1994).

[14] Fersht, A.R. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **7**, 3–9 (1997).

[15] Daggett, V. & Fersht, A. The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.* **4**, 497–502 (2003).

[16] Capaldi, A.P., Kleanthous, C. & Radford, S.E. Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol.* **9**, 209–216 (2002).

[17] Bycroft, M., Matouschek, A., Kellis, J.T., Serrano, L. & Fersht, A.R. Detection and characterization of a folding intermediate in barnase by NMR. *Nature* **346**, 488–490 (1990).

[18] Wetlaufer, D.B. Nucleation in protein folding–confusion of structure and process. *Trends Biochem. Sci.* **15**, 414–415 (1990).

[19] Clark, P. Protein folding in the cell: reshaping the folding funnel. *Trends Biochem. Sci.* **29**, 527–534 (2004).

[20] Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).

[21] Onuchic, J. Theory of protein folding. *Current Opinion in Structural Biology* (2004).

[22] Dinner, A.R., Sali, A., Smith, L.J., Dobson, C.M. & Karplus, M. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25**, 331–339 (2000).

[23] Ellis, R.J. & Minton, A.P. Protein aggregation in crowded environments. *Biol Chem* **387**, 485–497 (2006).

[24] DuBay, K.F., Pawar, A.P., Chiti, F., Zurdo, J., Dobson, C.M. & Vendruscolo, M. Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J Mol Biol* **341**, 1317–1326 (2004).

[25] Hashimoto, M., Rockenstein, E., Crews, L. & Masliah, E. Role of protein aggregation in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's diseases. *Neuromolecular Med* **4**, 21–36 (2003).

[26] Hipp, M.S., Park, S.H. & Hartl, F.U. Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.* (2014).

[27] Mirchev, R. & Ferrone, F.A. The structural link between polymerization and sickle cell disease. *J Mol Biol* **265**, 475–479 (1997).

[28] Zimmerman, S.B. & Trach, S.O. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J Mol Biol* **222**, 599–620 (1991).

[29] van den Berg, B., Ellis, R.J. & Dobson, C.M. Effects of macromolecular crowding on protein folding and aggregation. *Embo J* **18**, 6927–6933 (1999).

[30] Hong, J. & Gierasch, L.M. Macromolecular crowding remodels the energy landscape of a protein by favoring a more compact unfolded state. *J Am Chem Soc* **132**, 10445–10452 (2010).

[31] Wang, Y., He, H. & Li, S. Effect of Ficoll 70 on thermal stability and structure of creatine kinase. *Biochemistry Mosc.* **75**, 648–654 (2010).

[32] Hagen, S.J. Solvent viscosity and friction in protein folding dynamics. *Curr. Protein Pept. Sci.* **11**, 385–395 (2010).

[33] Plaxco, K.W., Simons, K.T. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**, 985–994 (1998).

[34] Dill, K.A., Fiebig, K.M. & Chan, H.S. Cooperativity in protein-folding kinetics. *P Natl Acad Sci USA* **90**, 1942–1946 (1993).

[35] Penkett, C.J., Redfield, C., Jones, J.A., Dodd, I., Hubbard, J., Smith, R.A., Smith, L.J. & Dobson, C.M. Structural and dynamical characterization of a biologically active unfolded fibronectin-binding protein from Staphylococcus aureus. *Biochemistry* **37**, 17054–17067 (1998).

[36] Matthews, J.M. & Sunde, M. Zinc fingers–folds for many occasions. *IUBMB Life* **54**, 351–355 (2002).

[37] Herbst, R., Schäfer, U. & Seckler, R. Equilibrium intermediates in the reversible unfolding of firefly (Photinus pyralis) luciferase. *J. Biol. Chem.* **272**, 7099–7105 (1997).

[38] Komar, A.A. A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.* **34**, 16–24 (2009).

[39] Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* **16**, 274–280 (2009).

[40] Preissler, S. & Deuerling, E. Ribosome-associated chaperones as key players in proteostasis. *Trends Biochem. Sci.* **37**, 274–283 (2012).

[41] Dunkle, J.A. & Cate, J.H.D. Ribosome structure and dynamics during translocation and termination. *Annu Rev Biophys* **39**, 227–244 (2010).

[42] Frank, J. & Gonzalez, R.L. Structure and dynamics of a processive Brownian motor: the translating ribosome. *Annu. Rev. Biochem.* **79**, 381–412 (2010).

[43] Selmer, M., Dunham, C.M., Murphy, F.V., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. & Ramakrishnan, V. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).

[44] Yusupov, M.M., Yusupova, G.Z., Baucom, A. & Lieberman, K. Crystal structure of the ribosome at 5.5 Å resolution. *Science* (2001).

[45] McCoy, A.J., Fucini, P., Noegel, A.A. & Stewart, M. Structural basis for dimerization of the Dictyostelium gelation factor (ABP120) rod. *Nat. Struct. Biol.* **6**, 836–841 (1999).

[46] Bocharov, E.V., Sobol, A.G., Pavlov, K.V., Korzhnev, D.M., Jaravine, V.A., Gudkov, A.T. & Arseniev, A.S. From structure and dynamics of protein L7/L12 to molecular switching in ribosome. *J. Biol. Chem.* **279**, 17697–17706 (2004).

[47] Voss, N.R., Gerstein, M., Steitz, T.A. & Moore, P.B. The geometry of the ribosomal polypeptide exit tunnel. *J Mol Biol* **360**, 893–906 (2006).

[48] Kramer, G., Boehringer, D., Ban, N. & Bukau, B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol* **16**, 589–597 (2009).

[49] Fedorov, A.N. & Baldwin, T.O. Cotranslational protein folding. *J. Biol. Chem.* **272**, 32715–32718 (1997).

[50] Kramer, G., Ramachandiran, V. & Hardesty, B. Cotranslational folding–omnia mea mecum porto? *Int. J. Biochem. Cell Biol.* **33**, 541–553 (2001).

[51] Zhang, G. & Ignatova, Z. Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Curr. Opin. Struct. Biol.* (2010).

[52] Seidelt, B., Innis, C.A., Wilson, D.N., Gartmann, M., Armache, J.P., Villa, E., Trabuco, L.G., Becker, T., Mielke, T., Schulten, K., Steitz, T.A. & Beckmann, R. Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* **326**, 1412–1415 (2009).

[53] Becker, T., Bhushan, S., Jarasch, A., Armache, J.P., Funes, S., Jossinet, F., Gumbart, J., Mielke, T., Berninghausen, O., Schulten, K., Westhof, E., Gilmore, R., Mandon, E.C. & Beckmann, R. Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* **326**, 1369–1373 (2009).

[54] Lu, J. & Deutsch, C. Folding zones inside the ribosomal exit tunnel. *Nat Struct Mol Biol* **12**, 1123–1129 (2005).

[55] Woolhead, C.A., McCormick, P.J. & Johnson, A.E. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116**, 725–736 (2004).

[56] Bhushan, S., Hoffmann, T., Seidelt, B., Frauenfeld, J., Mielke, T., Berninghausen, O., Wilson, D.N. & Beckmann, R. SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol.* **9**, e1000581 (2011).

[57] Kosolapov, A. & Deutsch, C. Tertiary interactions within the ribosomal exit tunnel. *Nat Struct Mol Biol* **16**, 405–411 (2009).

[58] Choi, K.M., Atkins, J.F., Gesteland, R.F. & Brimacombe, R. Flexibility of the nascent polypeptide chain within the ribosome–contacts from the peptide N-terminus to a specific region of the 30S subunit. *Eur. J. Biochem.* **255**, 409–413 (1998).

[59] Blobel, G. & Sabatini, D.D. Controlled proteolysis of nascent polypeptides in rat liver cell fractions. I. Location of the polypeptides within ribosomes. *J Cell Biol* **45**, 130–145 (1970).

[60] Nakatogawa, H. & Ito, K. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **108**, 629–636 (2002).

[61] Mitra, K., Schaffitzel, C., Fabiola, F., Chapman, M.S., Ban, N. & Frank, J. Elongation arrest by SecM via a cascade of ribosomal RNA rearrangements. *Mol. Cell* **22**, 533–543 (2006).

[62] Woolhead, C.A., Johnson, A.E. & Bernstein, H.D. Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol. Cell* **22**, 587–598 (2006).

[63] Lu, J. & Deutsch, C. Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J Mol Biol* **384**, 73–86 (2008).

[64] Hoffmann, A., Becker, A.H., Zachmann-Brand, B., Deuerling, E., Bukau, B. & Kramer, G. Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol. Cell* **48**, 63–74 (2012).

[65] Cabrita, L.D., Hsu, S.T.D., Launay, H., Dobson, C.M. & Christodoulou, J. Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *P Natl Acad Sci USA* **106**, 22239–22244 (2009).

[66] Siller, E., DeZwaan, D.C., Anderson, J.F., Freeman, B.C. & Barral, J.M. Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J Mol Biol* **396**, 1310–1318 (2010).

[67] Dobson, C.M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).

[68] Chen, W. & Helenius, A. Role of ribosome and translocon complex during folding of influenza hemagglutinin in the endoplasmic reticulum of living cells. *Mol. Biol. Cell* **11**, 765–772 (2000).

[69] Makeyev, E.V., Kolb, V.A. & Spirin, A.S. Enzymatic activity of the ribosome-bound nascent polypeptide. *FEBS Lett.* **378**, 166–170 (1996).

[70] Land, A., Zonneveld, D. & Braakman, I. Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. *The FASEB journal* (2003).

[71] Nicola, A.V., Chen, W. & Helenius, A. Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat Cell Biol* **1**, 341–345 (1999).

[72] Kudlicki, W., Chirgwin, J., Kramer, G. & Hardesty, B. Folding of an enzyme into an active conformation while bound as peptidyl-tRNA to the ribosome. *Biochemistry* **34**, 14284–14287 (1995).

[73] Frydman, J., Erdjument-Bromage, H., Tempst, P. & Hartl, F.U. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat. Struct. Biol.* **6**, 697–705 (1999).

[74] Evans, M.S., Sander, I.M. & Clark, P.L. Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. *J Mol Biol* **383**, 683–692 (2008).

[75] Kaiser, C.M., Goldman, D.H., Chodera, J.D., Tinoco, I. & Bustamante, C. The ribosome modulates nascent protein folding. *Science* **334**, 1723–1727 (2011).

[76] Kelkar, D.A., Khushoo, A., Yang, Z. & Skach, W.R. Kinetic analysis of ribosome-bound fluorescent proteins reveals an early, stable, cotranslational folding intermediate. *J. Biol. Chem.* **287**, 2568–2578 (2012).

[77] Fedorov, A.N. & Baldwin, T.O. Contribution of cotranslational folding to the rate of formation of native protein structure. *P Natl Acad Sci USA* **92**, 1227–1231 (1995).

[78] Netzer, W.J. & Hartl, F.U. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388**, 343–349 (1997).

[79] Kleizen, B., van Vlijmen, T. & de Jonge, H. Folding of CFTR is predominantly cotranslational. *Molecular cell* (2005).

[80] Khushoo, A., Yang, Z., Johnson, A.E. & Skach, W.R. Ligand-driven vectorial folding of ribosome-bound human CFTR NBD1. *Mol. Cell* **41**, 682–692 (2011).

[81] Ellis, J. Proteins as molecular chaperones. *Nature* **328**, 378–379 (1987).

[82] Barraclough, R. & Ellis, R. Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated intact pea chloroplasts. *Biochimica et biophysica acta* (1980).

[83] Specht, S., Miller, S.B.M., Mogk, A. & Bukau, B. Hsp42 is required for sequestration of protein aggregates into deposition sites in Saccharomyces cerevisiae. *J Cell Biol* **195**, 617–629 (2011).

[84] Martinez-Hackert, E. & Hendrickson, W.A. Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone. *Cell* **138**, 923–934 (2009).

[85] Karagoz, G.E., Duarte, A.M., Ippel, H., Uetrecht, C., Sinnige, T., van Rosmalen, M., Hausmann, J., Heck, A.J., Boelens, R. & Rudiger, S.G. N-terminal domain of human Hsp90 triggers binding to the cochaperone p23. *P Natl Acad Sci USA* **108**, 580–585 (2011).

[86] Hoffmann, A., Bukau, B. & Kramer, G. Structure and function of the molecular chaperone Trigger Factor. *Biochim. Biophys. Acta* **1803**, 650–661 (2010).

[87] Saibil, H. Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* **14**, 630–642 (2013).

[88] Kovermann, M., Schmid, F.X. & Balbach, J. Molecular function of the prolyl cis/trans isomerase and metallochaperone SlyD. *Biol Chem* **394**, 965–975 (2013).

[89] Castanié-Cornet, M.P., Bruel, N. & Genevaux, P. Chaperone networking facilitates protein targeting to the bacterial cytoplasmic membrane. *Biochim. Biophys. Acta* **1843**, 1442–1456 (2014).

[90] Dalbey, R.E., Kuhn, A., Zhu, L. & Kiefer, D. The membrane insertase YidC. *Biochim. Biophys. Acta* **1843**, 1489–1496 (2014).

[91] Goemans, C., Denoncin, K. & Collet, J.F. Folding mechanisms of periplasmic proteins. *Biochim. Biophys. Acta* **1843**, 1517–1528 (2014).

[92] Zückert, W.R. Secretion of Bacterial Lipoproteins: Through the Cytoplasmic Membrane, the Periplasm and Beyond. *Biochim. Biophys. Acta* **1843**, 1509–1516 (2014).

[93] Choi, D., Ryu, K.S. & Park, C. Structural alteration of Escherichia coli Hsp31 by thermal unfolding increases chaperone activity. *Biochim. Biophys. Acta* **1834**, 621–628 (2013).

[94] Kumsta, C. & Jakob, U. Redox-regulated chaperones. *Biochemistry* **48**, 4666–4676 (2009).

[95] Hong, W., Wu, Y.E., Fu, X. & Chang, Z. Chaperone-dependent mechanisms for acid resistance in enteric bacteria. *Trends Microbiol.* **20**, 328–335 (2012).

[96] Bepperling, A., Alte, F., Kriehuber, T., Braun, N., Weinkauf, S., Groll, M., Haslbeck, M. & Buchner, J. Alternative bacterial two-component small heat shock protein systems. *P Natl Acad Sci USA* **109**, 20407–20412 (2012).

[97] Street, T.O., Zeng, X., Pellarin, R., Bonomi, M., Sali, A., Kelly, M.J.S., Chu, F. & Agard, D.A. Elucidating the mechanism of substrate recognition by the bacterial Hsp90 molecular chaperone. *J Mol Biol* **426**, 2393–2404 (2014).

[98] Haslberger, T., Bukau, B. & Mogk, A. Towards a unifying mechanism for ClpB/Hsp104-mediated protein disaggregation and prion propagation. *Biochem. Cell Biol.* **88**, 63–75 (2010).

[99] Baker, T.A. & Sauer, R.T. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim. Biophys. Acta* **1823**, 15–28 (2012).

[100] Sawa, J., Heuck, A., Ehrmann, M. & Clausen, T. Molecular transformers in the cell: lessons learned from the DegP protease-chaperone. *Curr. Opin. Struct. Biol.* **20**, 253–258 (2010).

[101] Geibel, S. & Waksman, G. The molecular dissection of the chaperone-usher pathway. *Biochim. Biophys. Acta* **1843**, 1559–1567 (2014).

[102] Tokumoto, U. & Takahashi, Y. Genetic analysis of the isc operon in Escherichia coli involved in the biogenesis of cellular iron-sulfur proteins. *J. Biochem.* **130**, 63–71 (2001).

[103] Hartl, F. & Bracher, A. Molecular chaperones in protein folding and proteostasis. *Nature* (2011).

[104] Wandinger, S.K., Richter, K. & Buchner, J. The Hsp90 chaperone machinery. *J. Biol. Chem.* **283**, 18473–18477 (2008).

[105] Sharma, S.K., De los Rios, P., Christen, P., Lustig, A. & Goloubinoff, P. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nat. Chem. Biol.* **6**, 914–920 (2010).

[106] Zhou, J. & Xu, Z. The structural view of bacterial translocation-specific chaperone SecB: implications for function. *Mol. Microbiol.* **58**, 349–357 (2005).

[107] Sklar, J.G., Wu, T., Kahne, D. & Silhavy, T.J. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. *Genes Dev.* **21**, 2473–2484 (2007).

[108] Haslbeck, M., Franzmann, T., Weinfurtner, D. & Buchner, J. Some like it hot: the structure and function of small heat-shock proteins. *Nat Struct Mol Biol* **12**, 842–846 (2005).

[109] Neuwald, A.F., Aravind, L., Spouge, J.L. & Koonin, E.V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).

[110] Goloubinoff, P., Mogk, A., Zvi, A.P., Tomoyasu, T. & Bukau, B. Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. *P Natl Acad Sci USA* **96**, 13732–13737 (1999).

[111] Weibezahn, J., Schlieker, C., Tessarz, P., Mogk, A. & Bukau, B. Novel insights into the mechanism of chaperone-assisted protein disaggregation. *Biol Chem* **386**, 739–744 (2005).

[112] Seyffer, F., Kummer, E., Oguchi, Y., Winkler, J., Kumar, M., Zahn, R., Sourjik, V., Bukau, B. & Mogk, A. Hsp70 proteins bind Hsp100 regulatory M domains to activate AAA+ disaggregase at aggregate surfaces. *Nat Struct Mol Biol* **19**, 1347–1355 (2012).

[113] Joshi, S.A., Hersch, G.L., Baker, T.A. & Sauer, R.T. Communication between ClpX and ClpP during substrate processing and degradation. *Nat Struct Mol Biol* **11**, 404–411 (2004).

[114] Waksman, G. & Hultgren, S.J. Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nat Rev Microbiol* **7**, 765–774 (2009).

[115] Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A. & Xie, X.S. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).

[116] Tokuriki, N. & Tawfik, D.S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668–673 (2009).

[117] Rutherford, S.L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).

[118] Vorderwülbecke, S., Kramer, G., Merz, F., Kurz, T.A., Rauch, T., Zachmann-Brand, B., Bukau, B. & Deuerling, E. Low temperature or GroEL/ES overproduction permits growth of Escherichia coli cells lacking trigger factor and DnaK. *FEBS Lett.* **559**, 181–187 (2004).

[119] Genevaux, P., Keppel, F., Schwager, F., Langendijk-Genevaux, P.S., Hartl, F.U. & Georgopoulos, C. In vivo analysis of the overlapping functions of DnaK and trigger factor. *EMBO Rep.* **5**, 195–200 (2004).

[120] Deuerling, E., Patzelt, H., Vorderwülbecke, S., Rauch, T., Kramer, G., Schaffitzel, E., Mogk, A., Schulze-Specking, A., Langen, H. & Bukau, B. Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. *Mol. Microbiol.* **47**, 1317–1328 (2003).

[121] Teter, S.A., Houry, W.A., Ang, D., Tradler, T., Rockabrand, D., Fischer, G., Blum, P., Georgopoulos, C. & Hartl, F.U. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell* **97**, 755–765 (1999).

[122] Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A. & Bukau, B. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* **400**, 693–696 (1999).

[123] Stoller, G., Rucknagel, K.P., Nierhaus, K.H., Schmid, F.X., Fischer, G. & Rahfeld, J.U. A ribosome-associated peptidyl-prolyl cis/trans isomerase identified as the trigger factor. *Embo J* **14**, 4939–4948 (1995).

[124] Wiedmann, B., Sakai, H., Davis, T.A. & Wiedmann, M. A protein complex required for signal-sequence-specific sorting and translocation. *Nature* **370**, 434–440 (1994).

[125] Wang, L., Zhang, W., Wang, L., Zhang, X.C., Li, X. & Rao, Z. Crystal structures of NAC domains of human nascent polypeptide-associated complex (NAC) and its Œ±NAC subunit. *Protein Cell* **1**, 406–416 (2010).

[126] Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E. & Ban, N. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590–596 (2004).

[127] Kim, S.J., Jeong, D.G., Chi, S.W., Lee, J.S. & Ryu, S.E. Crystal structure of proteolytic fragments of the redox-sensitive Hsp33 with constitutive chaperone activity. *Nat. Struct. Biol.* **8**, 459–466 (2001).

[128] Bitto, E. & McKay, D.B. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10**, 1489–1498 (2002).

[129] Chen, H., Mustafi, S.M., LeMaster, D.M., Li, Z., Héroux, A., Li, H. & Hernández, G. Crystal structure and conformational flexibility of the unligated FK506-binding protein FKBP12.6. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 636–646 (2014).

[130] Saio, T., Guan, X., Rossi, P., Economou, A. & Kalodimos, C.G. Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science* **344**, 1250494 (2014).

[131] Lakshmipathy, S.K., Tomic, S., Kaiser, C.M., Chang, H.C., Genevaux, P., Georgopoulos, C., Barral, J.M., Johnson, A.E., Hartl, F.U. & Etchells, S.A. Identification of nascent chain interaction sites on trigger factor. *J. Biol. Chem.* **282**, 12186–12193 (2007).

[132] Merz, F., Boehringer, D., Schaffitzel, C., Preissler, S., Hoffmann, A., Maier, T., Rutkowska, A., Lozza, J., Ban, N., Bukau, B. & Deuerling, E. Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. *Embo J* **27**, 1622–1632 (2008).

[133] Zarnt, T., Tradler, T., Stoller, G., Scholz, C., Schmid, F.X. & Fischer, G. Modular structure of the trigger factor required for high activity in protein folding. *J Mol Biol* **271**, 827–837 (1997).

[134] Hesterkamp, T. & Bukau, B. Identification of the prolyl isomerase domain of Escherichia coli trigger factor. *FEBS Lett.* **385**, 67–71 (1996).

[135] Kristensen, O. & Gajhede, M. Chaperone binding at the ribosomal exit tunnel. *Structure* **11**, 1547–1556 (2003).

[136] Hesterkamp, T., Deuerling, E. & Bukau, B. The amino-terminal 118 amino acids of Escherichia coli trigger factor constitute a domain that is necessary and sufficient for binding to ribosomes. *J. Biol. Chem.* **272**, 21865–21871 (1997).

[137] Kramer, G., Rauch, T., Rist, W., Vorderwülbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E. & Bukau, B. L23 protein functions as a chaperone docking site on the ribosome. *Nature* **419**, 171–174 (2002).

[138] Kramer, G., Patzelt, H., Rauch, T., Kurz, T.A., Vorderwülbecke, S., Bukau, B. & Deuerling, E. Trigger factor peptidyl-prolyl cis/trans isomerase activity is not essential for the folding of cytosolic proteins in Escherichia coli. *J. Biol. Chem.* **279**, 14165–14170 (2004).

[139] Scholz, C., Stoller, G., Zarnt, T., Fischer, G. & Schmid, F.X. Cooperation of enzymatic and chaperone functions of trigger factor in the catalysis of protein folding. *Embo J* **16**, 54–58 (1997).

[140] Gupta, R., Lakshmipathy, S.K., Chang, H.C., Etchells, S.A. & Hartl, F.U. Trigger factor lacking the PPIase domain can enhance the folding of eukaryotic multi-domain proteins in Escherichia coli. *FEBS Lett.* **584**, 3620–3624 (2010).

[141] Guthrie, B. & Wickner, W. Trigger factor depletion or overproduction causes defective cell division but does not block protein export. *J Bacteriol* **172**, 5555–5562 (1990).

[142] Merz, F., Hoffmann, A., Rutkowska, A., Zachmann-Brand, B., Bukau, B. & Deuerling, E. The C-terminal domain of Escherichia coli trigger factor represents the central module of its chaperone activity. *J. Biol. Chem.* **281**, 31963–31971 (2006).

[143] Zeng, L.L., Yu, L., Li, Z.Y., Perrett, S. & Zhou, J.M. Effect of C-terminal truncation on the molecular chaperone function and dimerization of Escherichia coli trigger factor. *Biochimie* **88**, 613–619 (2006).

[144] Lakshmipathy, S.K., Gupta, R., Pinkert, S., Etchells, S.A. & Hartl, F.U. Versatility of trigger factor interactions with ribosome-nascent chain complexes. *J. Biol. Chem.* **285**, 27911–27923 (2010).

[145] Lill, R., Crooke, E., Guthrie, B. & Wickner, W. The "trigger factor cycle" includes ribosomes, presecretory proteins, and the plasma membrane. *Cell* **54**, 1013–1018 (1988).

[146] Baram, D., Pyetan, E., Sittner, A., Auerbach-Nevo, T., Bashan, A. & Yonath, A. Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. *P Natl Acad Sci USA* **102**, 12017–12022 (2005).

[147] Schlunzen, F., Wilson, D.N., Tian, P., Harms, J.M., McInnes, S.J., Hansen, H.A., Albrecht, R., Buerger, J., Wilbanks, S.M. & Fucini, P. The binding mode of the trigger factor on the ribosome: implications for protein folding and SRP interaction. *Structure* **13**, 1685–1694 (2005).

[148] Kaiser, C.M., Chang, H.C., Agashe, V.R., Lakshmipathy, S.K., Etchells, S.A., Hayer-Hartl, M., Hartl, F.U. & Barral, J.M. Real-time observation of trigger factor function on translating ribosomes. *Nature* **444**, 455–460 (2006).

[149] Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G., Weissman, J.S. & Bukau, B. Selective Ribosome Profiling Reveals the Cotranslational Chaperone Action of Trigger Factor In Vivo. *Cell* **147**, 1295–1308 (2011).

[150] Patzelt, H., Rüdiger, S., Brehmer, D., Kramer, G., Vorderwülbecke, S., Schaffitzel, E., Waitz, A., Hesterkamp, T., Dong, L., Schneider-Mergener, J., Bukau, B. & Deuerling, E. Binding specificity of Escherichia coli trigger factor. *P Natl Acad Sci USA* **98**, 14244–14249 (2001).

[151] Kerner, M.J., Naylor, D.J., Ishihama, Y., Maier, T., Chang, H.C., Stines, A.P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M. & Hartl, F.U. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* **122**, 209–220 (2005).

[152] Hoffmann, A., Merz, F., Rutkowska, A., Zachmann-Brand, B., Deuerling, E. & Bukau, B. Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *Journal of Biological Chemistry* **281**, 6539–6545 (2006).

[153] Tomic, S., Johnson, A.E., Hartl, F.U. & Etchells, S.A. Exploring the capacity of trigger factor to function as a shield for ribosome bound polypeptide chains. *FEBS Lett.* **580**, 72–76 (2006).

[154] Mashaghi, A., Kramer, G., Bechtluft, P., Zachmann-Brand, B., Driessen, A.J.M., Bukau, B. & Tans, S.J. Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* **500**, 98–101 (2013).

[155] Agashe, V.R., Guha, S., Chang, H.C., Genevaux, P., Hayer-Hartl, M., Stemp, M., Georgopoulos, C., Hartl, F.U. & Barral, J.M. Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. *Cell* **117**, 199–209 (2004).

[156] Ullers, R.S., Houben, E.N.G., Brunner, J., Oudega, B., Harms, N. & Luirink, J. Sequence-specific interactions of nascent Escherichia coli polypeptides with trigger factor and signal recognition particle. *J. Biol. Chem.* **281**, 13999–14005 (2006).

[157] Beck, K., Wu, L.F., Brunner, J. & Muller, M. Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. *Embo J* **19**, 134–143 (2000).

[158] Xie, J.B. & Zhou, J.M. Trigger factor assisted folding of green fluorescent protein. *Biochemistry* **47**, 348–357 (2008).

[159] Patzelt, H., Kramer, G., Rauch, T., Schonfeld, H.J., Bukau, B. & Deuerling, E. Three-state equilibrium of Escherichia coli trigger factor. *Biol Chem* **383**, 1611–1619 (2002).

[160] Maier, R., Eckert, B., Scholz, C., Lilie, H. & Schmid, F.X. Interaction of trigger factor with the ribosome. *J Mol Biol* **326**, 585–592 (2003).

[161] Rutkowska, A., Mayer, M.P., Hoffmann, A., Merz, F., Zachmann-Brand, B., Schaffitzel, C., Ban, N., Deuerling, E. & Bukau, B. Dynamics of trigger factor interaction with translating ribosomes. *J. Biol. Chem.* **283**, 4124–4132 (2008).

[162] Bornemann, T., Holtkamp, W. & Wintermeyer, W. Interplay between trigger factor and other protein biogenesis factors on the ribosome. *Nat Commun* **5**, 4180 (2014).

[163] Kleckner, I.R. & Foster, M.P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta* **1814**, 942–968 (2011).

[164] Neudecker, P., Lundström, P. & Kay, L.E. Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys. J.* **96**, 2045–2054 (2009).

[165] Boehr, D.D., Dyson, H.J. & Wright, P.E. An NMR perspective on enzyme dynamics. *Chem. Rev.* **106**, 3055–3079 (2006).

[166] Baldwin, A.J. & Kay, L.E. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **5**, 808–814 (2009).

[167] Zeeb, M. & Balbach, J. Protein folding studied by real-time NMR spectroscopy. *Methods* **34**, 65–74 (2004).

[168] Arrington, C.B. & Robertson, A.D. Correlated motions in native proteins from MS analysis of NH exchange: evidence for a manifold of unfolding reactions in ovomucoid third domain. *J Mol Biol* **300**, 221–232 (2000).

[169] Mittag, T., Schaffhausen, B. & Günther, U.L. Tracing kinetic intermediates during ligand binding. *J Am Chem Soc* **126**, 9017–9023 (2004).

[170] Vallurupalli, P. & Kay, L.E. Complementarity of ensemble and single-molecule measures of protein motion: a relaxation dispersion NMR study of an enzyme complex. *P Natl Acad Sci USA* **103**, 11910–11915 (2006).

[171] Vallurupalli, P., Scott, L., Hennig, M., Williamson, J.R. & Kay, L.E. New RNA labeling methods offer dramatic sensitivity enhancements in 2H NMR relaxation spectra. *J Am Chem Soc* **128**, 9346–9347 (2006).

[172] Tolman, J.R. & Ruan, K. NMR residual dipolar couplings as probes of biomolecular dynamics. *Chem. Rev.* **106**, 1720–1736 (2006).

[173] Lange, O.F., Lakomek, N.A., Farès, C., Schröder, G.F., Walter, K.F.A., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C. & de Groot, B.L. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475 (2008).

[174] Keramisanou, D., Biris, N., Gelis, I., Sianidis, G., Karamanou, S., Economou, A. & Kalodimos, C.G. Disorder-order folding transitions underlie catalysis in the helicase motor of SecA. *Nat Struct Mol Biol* **13**, 594–602 (2006).

[175] Fernandez, C. & Wider, G. TROSY in NMR studies of the structure and function of large biological macromolecules. *Current Opinion in Structural Biology* **13**, 570–580 (2003).

[176] Goto, N.K. & Kay, L.E. New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr. Opin. Struct. Biol.* **10**, 585–592 (2000).

[177] Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *P Natl Acad Sci USA* **94**, 12366–12371 (1997).

[178] Fiaux, J., Bertelsen, E.B., Horwich, A.L. & Wüthrich, K. NMR analysis of a 900K GroEL GroES complex. *Nature* **418**, 207–211 (2002).

[179] Tugarinov, V., Muhandiram, R., Ayed, A. & Kay, L.E. Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. *J Am Chem Soc* **124**, 10025–10035 (2002).

[180] Tugarinov, V. & Kay, L.E. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* **125**, 13868–13878 (2003).

[181] Tugarinov, V. & Kay, L.E. Quantitative NMR studies of high molecular weight proteins: application to domain orientation and ligand binding in the 723 residue enzyme malate synthase G. *J Mol Biol* **327**, 1121–1133 (2003).

[182] Korzhnev, D.M., Kloiber, K., Kanelis, V., Tugarinov, V. & Kay, L.E. Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. *J Am Chem Soc* **126**, 3964–3973 (2004).

[183] Christodoulou, J., Larsson, G., Fucini, P., Connell, S.R., Pertinhez, T.A., Hanson, C.L., Redfield, C., Nierhaus, K.H., Robinson, C.V., Schleucher, J. & Dobson, C.M. Heteronuclear NMR investigations of dynamic regions of intact Escherichia coli ribosomes. *P Natl Acad Sci USA* **101**, 10949–10954 (2004).

[184] Mulder, F.A.A., Bouakaz, L., Lundell, A., Venkataramana, M., Liljas, A., Akke, M. & Sanyal, S. Conformation and dynamics of ribosomal stalk protein L12 in solution and on the ribosome. *Biochemistry* **43**, 5930–5936 (2004).

[185] Ellis, J.P., Bakke, C.K., Kirchdoerfer, R.N., Jungbauer, L.M. & Cavagnero, S. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem. Biol.* **3**, 555–566 (2008).

[186] Hsu, S.T.D., Cabrita, L.D., Fucini, P., Dobson, C.M. & Christodoulou, J. Probing Protein Folding on the Ribosome by Solution State NMR Spectroscopy. *J Biomol Struct Dyn* **26**, 846–846 (2009).

[187] Eichmann, C., Preissler, S., Riek, R. & Deuerling, E. Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *P Natl Acad Sci USA* **107**, 9111–9116 (2010).

[188] Rutkowska, A., Beerbaum, M., Rajagopalan, N., Fiaux, J., Schmieder, P., Kramer, G., Oschkinat, H. & Bukau, B. Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies. *FEBS Lett.* **583**, 2407–2413 (2009).

[189] Keeler, J. *Understanding NMR Spectroscopy.* John Wiley & Sons (2011).

[190] Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. & Sykes, B.D. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* **6**, 135–140 (1995).

[191] Cavanagh, J., Fairbrother, W.J., Arthur G Palmer, I., Skelton, N.J. & Rance, M. *Protein NMR Spectroscopy.* Principles and Practice. Academic Press (2010).

[192] Ortega, A. & García de la Torre, J. Efficient, accurate calculation of rotational diffusion and NMR relaxation of globular proteins from atomic-level structures and approximate hydrodynamic calculations. *J Am Chem Soc* **127**, 12764–12765 (2005).

[193] Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. NMR Assignment and Secondary Structure Determination of an Octameric 110 kDa Protein Using TROSY in Triple Resonance Experiments. *J Am Chem Soc* **122**, 7543–7548 (2000).

[194] Kay, L.E. & Gardner, K.H. Solution NMR spectroscopy beyond 25 kDa. *Curr. Opin. Struct. Biol.* **7**, 722–731 (1997).

[195] Xu, Y., Zheng, Y., Fan, J.S. & Yang, D. A new strategy for structure determination of large proteins in solution without deuteration. *Nat. Methods* **3**, 931–937 (2006).

[196] Gardner, K.H. & Kay, L.E. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* **27**, 357–406 (1998).

[197] LeMaster, D. & Richards, F. NMR sequential assignment of Escherichia coli thioredoxin utilizing random fractional deuteriation. *Biochemistry* (1988).

[198] Grzesiek, S., Anglister, J. & Ren, H. Carbon-13 line narrowing by deuterium decoupling in deuterium/carbon-13/nitrogen-15 enriched proteins. Application to triple resonance 4D J connectivity of .... *Journal of the American ...* (1993).

[199] Pervushin, K. Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Q Rev Biophys* **33**, 161–197 (2000).

[200] Tugarinov, V., Kanelis, V. & Kay, L. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nature protocols* (2006).

[201] Nicholson, L.K., Kay, L.E., Baldisseri, D.M., Arango, J., Young, P.E., Bax, A. & Torchia, D.A. Dynamics of methyl groups in proteins as studied by proton-detected 13C NMR spectroscopy. Application to the leucine residues of staphylococcal nuclease. *Biochemistry* **31**, 5253–5263 (1992).

[202] Tugarinov, V. & Kay, L.E. Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. *Chembiochem* **6**, 1567–1577 (2005).

[203] Gardner, K.H., Rosen, M.K. & Kay, L.E. Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* **36**, 1389–1401 (1997).

[204] Metzler, W., Wittekind, M., Goldfarb, V. & Mueller, L. Incorporation of 1H/13C/15N-Ile,Leu,Val into a perdeuterated, 15N-labelled protein : potential in structure determination of large proteins by NMR. *Journal of the American ...* (1996).

[205] Sprangers, R., Velyvis, A. & Kay, L. Solution NMR of supramolecular complexes: providing new insights into function. *Nat. Methods* (2007).

[206] Gelis, I., Bonvin, A., Keramisanou, D. & Koukaki, M. Structural basis for signal sequence recognition by the 204-kDa translocase motor SecA determined by NMR. *Cell* (2007).

[207] Mittermaier, A.K. & Kay, L.E. Observing biological dynamics at atomic resolution using NMR. *Trends Biochem. Sci.* **34**, 601–611 (2009).

[208] Carr, H.Y. & Purcell, E.M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review* (1954).

[209] Meiboom, S. & Gill, D. Modified spinecho method for measuring nuclear relaxation times. *Review of scientific instruments* (1958).

[210] Carver, J.P. & Richards, R.E. A general two-site solution for the chemical exchange produced dependence of T 2 upon the carr-Purcell pulse separation. *Journal of Magnetic Resonance (1969)* (1972).

[211] Tugarinov, V., Sprangers, R. & Kay, L.E. Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. *J Am Chem Soc* **129**, 1743–1750 (2007).

[212] Bothe, J.R., Nikolova, E.N., Eichhorn, C.D., Chugh, J., Hansen, A.L. & Al-Hashimi, H.M. Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat. Methods* **8**, 919–931 (2011).

[213] Tugarinov, V. & Kay, L.E. Relaxation rates of degenerate 1H transitions in methyl groups of proteins as reporters of side-chain dynamics. *J Am Chem Soc* **128**, 7299–7308 (2006).

[214] Kalk, A. & Berendsen, H. Proton magnetic relaxation and spin diffusion in proteins. *Journal of Magnetic Resonance (1969)* (1976).

[215] Akasaka, K. Longitudinal relaxation of protons under cross saturation and spin diffusion. *Journal of Magnetic Resonance (1969)* (1981).

[216] Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. & Shimada, I. A novel NMR method for determining the interfaces of large protein-protein complexes. *Nat. Struct. Biol.* **7**, 220–223 (2000).

[217] Göbl, C., Madl, T., Simon, B. & Sattler, M. NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog Nucl Magn Reson Spectrosc* **80**, 26–63 (2014).

[218] Iwahara, J. & Schwieters, C.D. Ensemble approach for NMR structure refinement against 1H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a .... *Journal of the American ...* (2004).

[219] Solomon, I. & Bloembergen, N. Nuclear magnetic interactions in the HF molecule. *The journal of chemical physics* (1956).

[220] Peters, T. & Pinto, B.M. Structure and dynamics of oligosaccharides: NMR and modeling studies. *Curr. Opin. Struct. Biol.* **6**, 710–720 (1996).

[221] Battiste, J.L. & Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**, 5355–5365 (2000).

[222] Clore, G.M. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules .... *Chem. Rev.* (2009).

[223] Price, K.E., Lucas, L.H. & Larive, C.K. Analytical applications of NMR diffusion measurements. *Analytical and bioanalytical chemistry* (2004).

[224] Wilkins, D.K., Grimshaw, S.B., Receveur, V., Dobson, C.M., Jones, J.A. & Smith, L.J. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424–16431 (1999).

[225] Stejskal, E.O. & Tanner, J.E. Spin diffusion measurements: spin echoes in the presence of a timedependent field gradient. *The journal of chemical physics* (1965).

[226] Ferrage, F., Zoonens, M., Warschawski, D.E., Popot, J.L. & Bodenhausen, G. Slow diffusion of macromolecular assemblies by a new pulsed field gradient NMR method. *J Am Chem Soc* **125**, 2541–2545 (2003).

[227] Augustyniak, R., Ferrage, F., Damblon, C., Bodenhausen, G. & Pelupessy, P. Efficient determination of diffusion coefficients by monitoring transport during recovery delays in NMR. *Chem. Commun. (Camb.)* **48**, 5307–5309 (2012).

[228] Gasteiger, E., Hoogland, C. & Gattiker, A. Protein identification and analysis tools on the ExPASy server. *The proteomics ...* (2005).

[229] Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J Am Chem Soc* **127**, 8014–8015 (2005).

[230] Lescop, E., Schanda, P. & Brutscher, B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* **187**, 163–169 (2007).

[231] Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293 (1995).

[232] Goddard, T. *SPARKY 3*. University of California (2004).

[233] Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. & Laue, E.D. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).

[234] Kleckner, I.R. & Foster, M.P. GUARDD: user-friendly MATLAB software for rigorous analysis of CPMG RD NMR data. *J Biomol NMR* **52**, 11–22 (2012).

[235] Jeschke, G. DEER distance measurements on proteins. *Annual review of physical chemistry* (2012).

[236] Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H.W. Dead-time free measurement of dipole-dipole interactions between electron spins. *J. Magn. Reson.* **142**, 331–340 (2000).

[237] Jeschke, G., Chechik, V., Ionita, P. & Godt, A. DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Applied Magnetic ...* (2006).

[238] Ludlam, A.V., Moore, B.A. & Xu, Z. The crystal structure of ribosomal chaperone trigger factor from Vibrio cholerae. *P Natl Acad Sci USA* **101**, 13436–13441 (2004).

[239] Xu, Y. & Matthews, S. TROSY NMR spectroscopy of large soluble proteins. *Top Curr Chem* **335**, 97–119 (2013).

[240] Raine, A., Lovmar, M., Wikberg, J. & Ehrenberg, M. Trigger factor binding to ribosomes with nascent peptide chains of varying lengths and sequences. *J. Biol. Chem.* **281**, 28033–28038 (2006).

[241] Maier, R., Scholz, C. & Schmid, F.X. Dynamic association of trigger factor with protein substrates. *J Mol Biol* **314**, 1181–1190 (2001).

[242] Crooke, E., Guthrie, B., Lecker, S., Lill, R. & Wickner, W. ProOmpA is stabilized for membrane translocation by either purified E. coli trigger factor or canine signal recognition particle. *Cell* **54**, 1003–1011 (1988).

[243] Bremer, H. & Dennis, P.P. Modulation of chemical composition and other parameters of the cell by growth rate. In F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter & H.E. Umbarger, editors, *Escherichia coli and Salmonella: cellular and molecular biology*, pages 1553–1569. Wahington, D.C (1996).

[244] Dominguez, C., Boelens, R. & Bonvin, A.M.J.J. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731–1737 (2003).

[245] de Vries, S.J., van Dijk, A.D.J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T. & Bonvin, A.M.J.J. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* **69**, 726–733 (2007).

[246] Burré, J., Sharma, M., Tsetsenis, T. & Buchman, V. Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro. *Science* **329**, 1663–1667 (2010).

[247] Cookson, M.R., Xiromerisiou, G. & Singleton, A. How genetics research in Parkinson's disease is enhancing understanding of the common idiopathic forms of the disease. *Curr. Opin. Neurol.* **18**, 706–711 (2005).

[248] Chiti, F. & Dobson, C.M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366 (2006).

[249] Uversky, V.N., Li, J., Souillac, P., Millett, I.S., Doniach, S., Jakes, R., Goedert, M. & Fink, A.L. Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of alpha-synuclein assembly by beta- and gamma-synucleins. *Journal of Biological Chemistry* **277**, 11970–11978 (2002).

[250] Dedmon, M.M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M. & Dobson, C.M. Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J Am Chem Soc* **127**, 476–477 (2005).

[251] Allison, J.R., Varnai, P., Dobson, C.M. & Vendruscolo, M. Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. *J Am Chem Soc* **131**, 18314–18326 (2009).

[252] Davidson, W.S., Jonas, A., Clayton, D.F. & George, J.M. Stabilization of alpha-synuclein secondary structure upon binding to synthetic membranes. *Journal of Biological Chemistry* **273**, 9443–9449 (1998).

[253] Ulmer, T.S., Bax, A., Cole, N.B. & Nussbaum, R.L. Structure and dynamics of micelle-bound human alpha-synuclein. *Journal of Biological Chemistry* **280**, 9595–9603 (2005).

[254] Bertoncini, C.W., Jung, Y.S., Fernandez, C.O., Hoyer, W., Griesinger, C., Jovin, T.M. & Zweckstetter, M. Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *P Natl Acad Sci USA* **102**, 1430–1435 (2005).

[255] Rivers, R.C., Kumita, J.R., Tartaglia, G.G., Dedmon, M.M., Pawar, A., Vendruscolo, M., Dobson, C.M. & Christodoulou, J. Molecular determinants of the aggregation behavior of alpha- and beta-synuclein. *Protein Sci.* **17**, 887–898 (2008).

[256] Waudby, C.A., Camilloni, C., Fitzpatrick, A.W.P., Cabrita, L.D., Dobson, C.M., Vendruscolo, M. & Christodoulou, J. In-cell NMR characterization of the secondary structure populations of a disordered conformation of α-synuclein within E. coli cells. *PLoS ONE* **8**, e72286 (2013).

[257] Hsu, S.T.D., Cabrita, L.D., Fucini, P., Christodoulou, J. & Dobson, C.M. Probing Side-Chain Dynamics of a Ribosome-Bound Nascent Chain Using Methyl NMR Spectroscopy. *J Am Chem Soc* **131**, 8366–+ (2009).

[258] Hsu, S.T., Fucini, P., Cabrita, L.D., Launay, H., Dobson, C.M. & Christodoulou, J. Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *P Natl Acad Sci USA* **104**, 16516–16521 (2007).

[259] Dehouck, Y., Grosfils, A., Folch, B., Gilis, D., Bogaerts, P. & Rooman, M. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* **25**, 2537–2543 (2009).

[260] Matlack, K.E. & Walter, P. The 70 carboxyl-terminal amino acids of nascent secretory proteins are protected from proteolysis by the ribosome and the protein translocation apparatus of the endoplasmic reticulum membrane. *J. Biol. Chem.* **270**, 6170–6180 (1995).

[261] Felitsky, D.J., Lietzow, M.A. & Dyson, H.J. Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. In *Proceedings of the ...* (2008).